



HAL
open science

Reconnaissance de scénario par les Modèles de Markov Cachés Crédibilistes : Application à l'interprétation automatique de séquences vidéos médicales

Arnaud Ahouandjinou

► **To cite this version:**

Arnaud Ahouandjinou. Reconnaissance de scénario par les Modèles de Markov Cachés Crédibilistes : Application à l'interprétation automatique de séquences vidéos médicales. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université du Littoral Côte d'Opale; Université d'Abomey-Calavi (Bénin), 2014. Français. NNT : 2014DUNK0380 . tel-01292584

HAL Id: tel-01292584

<https://theses.hal.science/tel-01292584v1>

Submitted on 5 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : (à ne mettre qu'après la délivrance du doctorat)

ECOLE DOCTORALE SPI DE L'UNIVERSITE LILLE NORD-DE-FRANCE
&
ECOLE DOCTORALE MIPT DE L'UNIVERSITE D'ABOMEY-CALAVI DU BENIN

THESE

présentée en vue
d'obtenir le grade de

DOCTEUR

en
Spécialité : Génie Informatique, Automatique et Traitement du Signal
par

Arnaud AHOANDJINO

DOCTORAT DELIVRE CONJOINTEMENT
PAR L'UNIVERSITE DU LITTORAL CÔTE D'OPALE ET L'UNIVERSITE D'ABOMEY-CALAVI

Titre de la thèse :

Reconnaissance de scénario par les Modèles de Markov Cachés Crédibilistes. Application à l'interprétation automatique de séquences vidéo médicales

Soutenue le Décembre 2014 devant le jury d'examen :

Président	<i>(le nom du président ne peut être arrêté que le jour de la soutenance)</i>
Rapporteur	<i>Antonio, PINTI, Maître de conférences HDR, Université d'Orléans</i>
Rapporteur	<i>Kokou ASSOGBA, Maître de conférences HDR, UAC</i>
Membre	<i>Hippolyte AGBOTON, Professeur titulaire, UAC</i>
Membre	<i>Pierre GOUTON, Professeur titulaire, Université de Bourgogne</i>
Directeur de thèse	<i>Cina MOTAMED, Maître de conférences HDR, ULCO</i>
Co-directeur de thèse	<i>Eugène C. EZIN, Maître de conférences, UAC</i>

Thèse préparée dans le Laboratoire d'Informatique, Signal et Image de la Côte d'Opale (LISIC)
Université du Littoral Côte d'Opale, 50 rue F. Buisson –BP 719, 62228 Calais Cedex, France.

&

L'Unité de Recherche d'Informatique et Sciences Appliquées (URISA)
Université d'Abomey-Calavi-IMSP, BP 613 Porto-Novo, Bénin.

Remerciements

A mes directeurs de thèses M. Cina MOTAMED et M. Eugène C. EZIN pour l'encadrement sans faille, leurs précieux conseils et surtout leurs exceptionnelles qualités humaines, je dis très sincèrement merci.

Au professeur Joël TOSSA, Directeur de l'Institut de Mathématiques et de Sciences Physiques (IMSP) pour son accompagnement et sa disponibilité pour les services administratifs. De même, au professeur Pierre GOUTON qui m'a permis de faire mes premiers pas significatifs dans le monde de la recherche par son soutien indéfectible et son accompagnement technique pour ma première publication, j'en suis infiniment très reconnaissant.

A mes collègues de bureau de l'ESGIS avec qui je forme vraiment une famille et avec qui je travaille en synergie malgré nos différences. Tout cela n'a été possible que grâce au chef département ingénierie informatique et réseau Dr. Bethel ATOHOUN, un homme tout simplement exceptionnel. Par la même occasion, j'adresse mes vifs remerciements au directeur général de l'ESGIS qui m'a permis pendant toutes ces années de financer ma thèse.

J'adresse mes sincères remerciements aux rapporteurs de ma thèse, dont les remarques et suggestions permettront d'améliorer la qualité de ce manuscrit. Je remercie par ailleurs tous les membres de mon jury pour leur disponibilité afin d'évaluer ce travail. En particulier, j'adresse toute ma gratitude au professeur Hippolyte AGBOTON qui m'a permis de déployer le système de vidéosurveillance expérimental et de collecter la base de données vidéos médicales aux unités de soins intensifs du service cardiologique du Centre National Hospitalier Universitaire (CNHU-HKM) du Bénin.

Mes vifs et chaleureux remerciements vont aussi à l'endroit de ma famille surtout mon père qui a consenti tant d'effort matériel et moral pour me soutenir pendant toutes ces longues années d'études. Je n'oublierai pas ma mère pour son soutien moral indéfectible et ses prières sans relâche pour ma réussite professionnelle. Mon fils, sa mère pour leur accompagnement pendant les moments difficiles.

A tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail, je leur témoigne ma reconnaissance.

Table des matières

Remerciements.....	ii
Table des matières.....	iii
Liste des Figures	vii
Liste des Tableaux	ix
Introduction générale	1

Chapitre 1 : *Des voies vers un système de vidéosurveillance multicaméras en Unités de Soins Intensifs (USIs)*

1.1.	Les systèmes de vidéosurveillance intelligents.....	7
1.1.1.	Les principales fonctionnalités en vision par ordinateur	8
1.1.2.	Une pluralité d'applications dans divers domaines.....	10
1.1.2.1.	Une solution de sécurisation des biens et des personnes.....	10
1.1.2.2.	Une solution de monitoring d'activités sportives.....	11
1.1.2.3.	Une solution de monitoring d'activités humaines en milieu médical	13
1.1.3.	Bilan des spécificités liées à chaque type d'applications.....	15
1.2.	Présentation d'une Unité de Soins Intensifs	18
1.2.1.	Structure organisationnelle et composantes d'une USI.....	18
1.2.2.	Caractéristiques et spécificités des données et des connaissances d'une USI médicale	21
1.2.2.1.	La surveillance des patients en unités de soins intensifs.....	22
1.2.2.2.	Caractéristiques des données en unités de soins intensifs.....	24
1.3.	Le système de vidéo-surveillance multicaméras proposé en complément à l'existant en USIs	27
1.3.1.	Architecture physique et organisation multicaméras du système de vidéo surveillance en USIs.....	29
1.3.2.	Architecture logique du système de vidéo surveillance en USIs	32
1.4.	Vers une Boîte Noire Médicale en USIs	32

Chapitre 2 : *Etat de l'art des traitements de base d'un système de vision pour la reconnaissance de scénario*

2.1.	L'acquisition et le prétraitement d'images	37
2.1.1.	L'acquisition de séquences d'images vidéo	37
2.1.2.	Calibrage et paramètres d'une caméra.....	38
2.1.3.	Le prétraitement des images vidéos acquises	40
2.1.3.1.	Les corrections photométriques.....	41
2.1.3.2.	Les corrections géométriques.....	42

2.2.	Les traitements bas niveaux	44
2.2.1.	La détection d'objets mobiles dans une vidéo	45
2.2.1.1.	Technique basée sur la différence entre deux images consécutives.....	46
2.2.1.2.	Technique basée sur la soustraction de fond	47
2.2.1.3.	Technique basée sur le flot optique	48
2.2.2.	Le suivi d'objets détectés mobiles dans une vidéo	49
2.2.2.1.	Techniques déterministes de suivi d'objets mobiles.....	50
2.2.2.2.	Techniques probabilistes Bayésiennes ou non de suivi d'objets	51
2.2.3.	Bilan de détection et du suivi pour l'approche adoptée	52
2.3.	Les traitements haut niveau.....	55
2.3.1.	Approches orientées méthodes symboliques	55
2.3.1.1.	Les méthodes symboliques sans contrainte temporelle.....	56
2.3.1.2.	Les méthodes symboliques avec contraintes temporelles	58
2.3.1.3.	Limites des méthodes symboliques	60
2.3.2.	Approches orientées méthodes numériques	61
2.3.2.1.	Les réseaux de neurones	64
2.3.2.2.	Les réseaux bayésiens dynamiques	65
2.3.2.3.	Les chaînes de Markov cachées et extensions.....	67
2.3.3.	Bilan des techniques de reconnaissance de scénario	70
2.4.	Module de gestion d'incertitude	72
2.4.1.	Cadre de raisonnement probabiliste	73
2.4.2.	Cadre de raisonnement possibiliste	73
2.4.3.	Cadre de raisonnement crédibiliste ou évidentielle	74
2.4.4.	Bilan et choix du cadre de raisonnement pour la gestion de l'incertain	75

Chapitre 3 : *Reconnaissance de scénario par les Modèles de Markov Cachés Hiérarchiques et Temporels*

3.1	Les modèles de Markov cachés probabilistes	79
3.1.1.	Présentation du modèle sur un cas d'école : balles dans des urnes	79
3.1.1.1.	Définition et composantes du modèle	80
3.1.1.2.	Problèmes de base et mécanismes de résolution	81
3.1.2.	Typologie et topologie des modèles de Markov	83
3.2.	Les extensions du modèle standard.....	86
3.3.	Les Modèles de Markov Cachés Hiérarchiques et Temporels.....	90
3.3.1.	Modèle de Markov Caché Hiérarchique (MMCH).....	90
3.3.1.1.	Définition et topologie d'un MMCH	91
3.3.1.2.	Modélisation et composantes du modèle MMCH.....	91
3.3.2.	Modèle Semi-Markovien Caché avec durée d'état explicite.....	93

3.3.3.	Modélisation de la durée d'état dans un MMCHT	94
3.3.3.1.	Distribution Discrète pour la durée d'état d'un MMCH	94
3.3.3.2.	Distribution continue pour la durée d'état d'un MMCHT	95
3.3.4.	Mécanismes de ré-estimation des paramètres d'un MMCH	96
3.3.5.	Mécanismes d'apprentissage et de reconnaissance de scénarios par exploitation d'un MMCHT	97
3.4.	Application à la reconnaissance des scénarios « <i>utiliser un distributeur</i> » et « <i>bagage abandonné</i> »	99
3.5.	Apports du modèle MMCHT	107
<hr/>		
Chapitre 4 : Reconnaissance de scénario par les Modèles de Markov Cachés Hiérarchico-Temporels et Crédibilistes(MMCHTC)		
<hr/>		
4.1.	Outils de gestion d'incertitude type crédibiliste	110
4.1.1.	De la théorie de l'évidence aux modèles des croyances transférables	110
4.1.2.	Formalismes et fonctions de croyances du MCT	111
4.1.2.1.	Formalisme et cadre de raisonnement	111
4.1.2.2.	Définitions et fonctions de croyances	113
4.2.	Opérations de bases dans le cadre du MCT	117
4.2.1.	Le PMI : principe du minimum d'information	118
4.2.2.	Règles de base de combinaisons d'informations	119
4.2.2.1.	Combinaison d'informations issues de sources distinctes et fiables	119
4.2.2.2.	Combinaison d'informations issues de sources distinctes et non fiables	121
4.2.3.	Prise de décision à partir de fonctions de croyance	122
4.2.3.1.	Prise de décision à partir de la probabilité pignistique.....	122
4.2.3.2.	Prise de décision à partir de la fonction plausibilité.....	123
4.2.4.	Gestion du conflit.....	123
4.3.	Mécanismes pour la reconnaissance de scénario par exploitation des modèles de Markov cachés hiérarchico-temporels crédibilistes	124
4.3.1.	Modèles de Markov cachés Crédibilistes.....	125
4.3.2.	Gestion de la temporalité des séquences et de la hiérarchie des scénarios par le modèle des MMC crédibiliste	127
4.3.3.	Apprentissage des paramètres du modèle	129
4.3.4.	Reconnaissance de scénarios crédibilistes	131
4.3.4.1.	Propagation <i>forward</i> crédibiliste	131
4.3.4.2.	Propagation <i>backward</i> crédibiliste et temporelle	133
4.4.	Application à la reconnaissance des scénarios d'actes « <i>potentiellement criminel</i> » et « <i>criminel</i> »	134
4.5.	Apport du MCT pour la gestion d'imperfections.....	138

Chapitre 5 : Reconnaissance d'activités dans des vidéos médicales et expérimentations

5.1.	Contexte expérimental	142
5.1.1.	Vue d'ensemble du système expérimental	142
5.1.2.	Description du système visuel expérimental proposé pour le monitoring des patients en USIs	144
5.1.3.	Description de la base de vidéos médicales d'USIs exploitées	145
5.1.3.1.	Caractérisation de la base de vidéos médicales	145
5.1.3.2.	Description et modélisation des scénarios à reconnaître	146
5.2.	Prétraitement des images médicales d'USIs	150
5.2.1.	Modèle de prise en compte des distorsions	150
5.2.2.	Approche de correction multi-vue	151
5.2.2.1.	Principe de correction de distorsion multi-vue	151
5.2.2.2.	Algorithme de correction de distorsion multi-vue	153
5.2.2.3.	Expérimentations et résultats	155
5.3.	Détection et suivi des objets dans les scènes médicales	157
5.3.1.	Détection des objets en mouvement de la scène médicale avec des indicateurs crédibilistes	158
5.3.2.	Suivi des objets en mouvement de la scène médicale avec des indicateurs crédibilistes	161
5.4.	Reconnaissance de scénarios dans des vidéos de scènes médicales	164
5.4.1.	Extraction de caractéristiques et construction du codebook	165
5.4.2.	Reconnaissance de scénarios probabilistes et crédibilistes par les modèles de Markov cachés hiérarchico-temporels	166
5.4.2.1.	Reconnaissance avec un MMC probabiliste intégrant une contrainte hiérarchique et temporelle	166
5.4.2.2.	Reconnaissance avec un MMC crédibiliste intégrant une contrainte temporelle et hiérarchique	173
5.4.3.	Résultats expérimentaux et analyse	178
Annexe A	Présentation des outils exploités pour le prétraitement	x
Annexe B	Présentation des mécanismes algorithmiques des MMCS utilisés	xv
Annexe C	Présentation des concepts et des mécanismes de base du Modèle de Croyances Transférables à travers un exemple didactique	xxi
Annexe D	Liste des publications	xxiv
Bibliographie	xxvi
Résumé	xli

Liste des Figures

Figure 1.1.	Illustration des principales technologies de vidéosurveillance multicaméras. <i>Les flèches indiquent la circulation du flux d'information entre les différents modules</i> [Wan12].	8
Figure 1.2.	Illustration de l'organisation type d'une salle de soins intensifs [SAS04].	21
Figure 1.3.	Architecture du système de traitement de vidéosurveillance multicaméras.	29
Figure 1.4.	: Topologie physique du système de vidéosurveillance multicaméras [AME13].	31
Figure 2.1.	Architecture en couche modulaire de base d'un système de vision pour la reconnaissance de scénario dans une séquence vidéo [AME14].	36
Figure 2.2.	Modèle géométrique sténopé d'une caméra [AME12].	38
Figure 2.3.	Modélisation du graphe d'état [TR99].	57
Figure 2.4.	Détection d'une chute avec trajectoire segmentée [NCM04].	64
Figure 2.5.	Exemple de représentation d'un réseau Bayésien dynamique [Zia10].	67
Figure 3.1.	Architecture d'un modèle de Markov caché à trois états.	81
Figure 3.2.	Architecture d'un modèle de Markov caché à trois états	81
Figure 3.3.	Exemple d'un modèle de Markov caché continu à trois états.	84
Figure 3.4.	Exemple d'un modèle de Markov caché discret à trois états.	84
Figure 3.5.	Exemple d'un modèle de Markov caché ergodique à trois états.	85
Figure 3.6.	Exemple de topologie gauche-droite séquentielle d'un MMC à trois états.	86
Figure 3.7.	Exemple la topologie gauche-droite d'un MMC parallèle à six états.	86
Figure 3.8.	: Exemple d'un modèle de Markov caché Hiérarchique.	91
Figure 3.9.	Exemple d'un MMCH à trois niveaux [Zia10].	93
Figure 3.10.	Structure hiérarchique du MMCHT exploité pour reconnaître le scénario « <i>utiliser un distributeur</i> ».	100
Figure 3.11.	Reconnaissance d'événements « <i>utiliser un distributeur</i> » dans des séquences vidéo tests de la base de données PETS'04.	102
Figure 3.12.	Reconnaissance d'événements « <i>utiliser un distributeur</i> » dans des séquences vidéo tests de la base de données PETS'04.	103
Figure 3.13.	Taux de reconnaissance d'événements « <i>utiliser un distributeur</i> » dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution de Poisson pour la modélisation du temps de séjour.	103
Figure 3.14.	Taux de reconnaissance d'événements « <i>utiliser un distributeur</i> » dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution gaussienne pour la modélisation du temps de séjour.	104
Figure 3.15.	Taux de reconnaissance d'événements « <i>abandon de bagage</i> » dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution gaussienne pour la modélisation du temps de séjour.	105
Figure 3.16.	Taux de reconnaissance d'événements Reconnaissance d'événements « <i>bagage abandonné</i> » dans des séquences vidéo tests de la base de données PETS'06.	105
Figure 3.17.	Reconnaissance d'événements « <i>abandon de bagage</i> » les graphes (a) et (b) correspond aux résultats de la probabilité de reconnaissance en fonction du numéro de la séquence de la base test PETS'04.	106
Figure 4.1.	Représentation abstraite des mécanismes dans le Modèle des Croyances Transférables [Ram07].	112
Figure 4.2.	Illustration du processus de fusion d'informations multi sources [Ram07].	113
Figure 4.3.	Illustration d'un processus d'hiérarchisation à trois niveaux [Mer06].	128
Figure 4.4.	Phases de calcul pour de mise en œuvre de l'algorithme de Baum-Welch	129
Figure 4.5.	Modèle graphique représentant les propagations forward et backward crédibilistes entre deux instants [Ram07].	131

Figure 4.6. Illustration du contexte de surveillance multicaméras (4 caméras) proposé par la base de données test PETS 2014. Cas de scénario de comportement « <i>potentiellement criminel</i> »...	135
Figure 4.7. Illustration du contexte de surveillance multicaméras (4 caméras) proposé par la base de données test PETS 2014. Cas de scénario de comportement « <i>criminel</i> ».	135
Figure 4.8. Reconnaissance d'événements de comportements « <i>potentiellement criminel</i> » dans des séquences vidéo tests de la base de données PETS' 14.	136
Figure 4.9. Reconnaissance d'événements de comportements « <i>criminel</i> » dans des séquences vidéo tests de la base de données PETS' 14.	137
Figure 5.1. Vue en plan côté de la salle de soins intensifs du système expérimental.	143
Figure 5.2. Illustration du calibrage de système de vision dans la salle de soins d'USIs.	143
Figure 5.3. Architecture globale de chaîne de traitement du système de reconnaissance.	145
Figure 5.4. Exemples d'images intégrant des scénarios reconnus par notre système.	146
Figure 5.5. Illustration d'exemples de séquences d'images collectées par caméra.	146
Figure 5.6. Organigramme de correction de distorsion multi-vue [AME ⁺ 12].	152
Figure 5.7. Illustration d'exemple d'une paire d'images stéréoscopiques.	155
Figure 5.8. Tracé des lignes épipolaires d'une paire d'images stéréoscopiques	155
Figure 5.9. Illustration d'exemples d'images distordues corrigées.	156
Figure 5.10. Illustration d'exemple d'images distordues corrigées.	156
Figure 5.11. Organisation globale de l'algorithme de détection [Zia10].	160
Figure 5.12. Modélisation de la salle soins à base de régions d'intérêts [AME13].	162
Figure 5.13. Illustration de deux points de vue différents de la scène observée respectivement de la caméra 4 et 2 à un instant t .	163
Figure 5.14. Illustration de deux points de vue différents de la scène observée respectivement de la caméra 4 et 2 à un instant $t+k$.	163
Figure 5.15. Résultats de détection et de suivi d'objets mobiles pour divers cas de scénario à reconnaître relatif à l'objet médecin.	164
Figure 5.16. Séquence des symboles du codebook par quantification vectorielle à partir d'un exemple de trajectoire [Zia10].	165
Figure 5.17. construction d'un codebook à partir de région d'intérêts définis [AME14b].	165
Figure 5.18. Structure et format des régions d'intérêts pour le codebook à base d'une image masque [AME14b].	166
Figure 5.19. Modélisation du scénario «le médecin visite ou fait des soins médicaux à un patient d'un lit n ».	169
Figure 5.20. Fréquence de la durée passée par le MMC dans les événements élémentaires du scénario 1.	170
Figure 5.21. Estimation statistique de la durée passée par le modèle dans l'un des états du scénario 1.	171
Figure 5.22. Estimation statistique de la durée passée par le modèle dans l'un des états du scénario 1.	171
Figure 5.23. Illustration de reconnaissance de scénario liée aux activités du médecin.	172
Figure 5.24. Topologie du MMC permettant de reconnaître le scénario « <i>le médecin fait des soins médicaux au patient du lit 1</i> » ou « <i>le médecin visite le patient hospitalisé au lit 1</i> ».	173
Figure 5.25. Topologie du MMC permettant de reconnaître le scénario « <i>le médecin fait un tour de visite dans toutes les salles de soins</i> ».	173
Figure 5.26. Topologie du MMC permettant de reconnaître le scénario « <i>le patient s'agite</i> » et le scénario « <i>le patient ressent des douleurs</i> ».	178
Figure 5.27. Topologie du MMCHT permettant de reconnaître le scénario « <i>le patient convulse</i> »	178
Figure 5.28. Figure Illustration de reconnaissance de scénario liée aux activités du médecin.	181
Figure 5.29. Quelques résultats de reconnaissance de scénarios « <i>le patient s'agite ou ressent des douleurs</i> ».	183

Figure 5.30. Histogramme du taux de reconnaissance des différentes MMCs avec une distribution de Poisson (%). 184
Figure 5.31. Histogramme du taux de reconnaissance des différentes MMCs avec une distribution de Normale (%). 184

Liste des Tableaux

Tableau 1.1. Tableau comparatif des caractéristiques et spécificités de quelques applications de vidéosurveillance intelligente selon le domaine d'application. 17
Tableau 2.1. Tableau comparatif de quelques méthodes de détection de mouvement et de suivi d'objets mobiles dans des séquences vidéo. 54
Tableau 2.2. Illustration de modélisation du graphe d'état [TR99]. 57
Tableau 2.3. Illustration de quelques exemples de mouvement du corps et définitions de comportement pour la base de données de geste [HJB02]. 65
Tableau 2.4. Bilan comparatif des approches de reconnaissance automatique d'activité. 71
Tableau 4.1. Les fonctions de croyance et leur utilité. 116
Tableau 4.2. Résultat de la transformation d'une fonction masse m donnée en *plausibilité* (pl), *crédibilité* (Cr), *implacabilité* (b), et *communalité* (q). 117
Tableau 4.3. Présentation de la correspondance entre notations probabilistes et celles crédibilistes. 127
Tableau 5.1. Description des différentes classes de scénarios reconnues. 149
Tableau 5.2. Taux de reconnaissance des différentes MMCs avec une distribution de Poisson (%). 183
Tableau 5.3. Taux de reconnaissance des différentes MMCs avec une distribution Gaussienne (%). 183

Introduction générale

La vidéosurveillance a connu un attrait significatif dans de nombreux domaines technologiques à travers des applications de recherche en vision par ordinateur et en intelligence artificielle. L'une des premières utilisations d'un système de surveillance remonte dans les années 1930 et 1940 dans le cadre d'une application monitoring et contrôle en milieu industriel [WTK⁺11, AJC11]. Les systèmes traditionnels de vidéosurveillance communément appelés *Close-Circuit TeleVision* (CCTV) ont rapidement évolué vers des systèmes de vidéosurveillance automatisés qui exploitent des techniques plus efficaces de la vision par ordinateur et intègre le temps réel. Ces systèmes de vidéosurveillances intelligents sont basés sur des techniques d'analyse et d'interprétation automatique vidéo et assistent le personnel de sécurité ou de surveillance dans l'exercice de leurs fonctions [CFA13].

1. Contexte et motivation

L'analyse de l'activité humaine dans des séquences vidéo est devenue un axe de recherche très actuel et reste en cours d'investigation dans la communauté de vision par ordinateur. Au cours de cette dernière décennie, elle a connu une évolution considérable qui se traduit par le développement de nombreuses applications qui sont bon marché dans divers domaines [AJC11, BXL11, LJK11]. Parmi ces domaines, nous avons la surveillance des lieux publics comme une gare, les aéroports, les hypermarchés, les sites militaires et la surveillance en environnement industriel. Plus récemment, plusieurs applications de type médical [RMSA⁺11, ZBT10, RCC⁺13] commencent à émerger. C'est dans ce cadre que s'inscrit le domaine d'application visé par nos travaux de thèse, qui reste nouveau dans le cadre d'une solution de vidéosurveillance intelligente en environnement hospitalier. La surveillance en milieu médical et spécifiquement en Unité de Soins Intensifs (USI) est d'un enjeu capital pour assister et maintenir les fonctions vitales des patients qui sont dans un état sanitaire instable mais ne pouvant bénéficier en permanence de la présence effective du corps médical limité et très sollicité. Dans un milieu aussi complexe et sensible, de nombreuses études cliniques, sociologiques et économiques ont prouvé l'opportunité et

la nécessité de développer des systèmes d'aide à la décision médicale [SAS04]. Cependant, la décision finale est laissée au médecin ou au praticien hospitalier et plus généralement à l'utilisateur du système. Face aux problèmes (des alarmes moins informatives, taux de fausses alarmes très élevé, l'incomplétude des connaissances médicales) récurrents et bien connus des systèmes de surveillance actuels en unités de soins intensifs médicaux [Guy07, SGM07], il est nécessaire d'améliorer les performances de la surveillance en USI.

Les travaux présentés dans cette thèse concernent le développement d'une solution de vidéo surveillance intelligente pour coupler le système de monitoring existant en USI et qui a pour principales fonctionnalités [AME14a] :

- La détection intelligente d'alarmes visuelles plus informatives que celles sonores du système existant afin d'aider le personnel médical ;
- La sélection intelligente de vidéo basée sur la reconnaissance automatique d'événements importants.

Doter le système de surveillance actuel dans les unités de soins intensifs hospitaliers de l'ensemble des fonctionnalités ci-dessus via le développement d'un système de vidéo surveillance intelligente constitue une amélioration notable de la qualité de service (QoS) médical en USIs [AME14a].

2. Contributions

Les travaux de recherche réalisés dans cette thèse ont pour objet l'étude et la mise en œuvre d'un système de vidéo surveillance intelligent. Plus spécifiquement, la conception d'un système de vidéosurveillance automatique et intelligent en milieu hospitalier (unité de soins intensifs) qui est un domaine d'application difficile du fait de la sensibilité du secteur médical car relatif à la santé humaine.

Dans un premier temps, nous avons abordé les différentes étapes d'un système de vidéosurveillance intelligent en vision par ordinateur pour l'analyse et l'interprétation de séquences vidéo. Les contributions majeures et originales de ces travaux se situent principalement dans la couche haute du système d'interprétation pour la reconnaissance crédibiliste d'événements vidéo. L'architecture globale du des traitements de vision du système est organisée en module de couches hiérarchiques. La première couche, concerne d'une part, les prétraitements pour lesquels nous avons

proposé une nouvelle technique élégante et robuste. Elles concernent aussi, le choix de l'approche de détection et de suivi d'objets dans les séquences d'images qui est basé sur un critère de temps réel et d'adaptation aux algorithmes des couches supérieures d'interprétation. Notre apport à ce niveau de traitement a été d'intégrer des métriques qualitatives qui permettent d'évaluer la qualité des résultats obtenus par ces traitements de vision. Ces métriques qualitatives sont exploitées pour propager l'incertitude relative aux traitements de vision bas niveau à travers les autres couches de traitements haut niveau afin de qualifier et d'améliorer les traitements des couches hautes.

Dans un second temps, nous avons abordé la problématique de reconnaissance de scénario en orientant notre choix vers des techniques utilisées en vision par ordinateur. Nous avons opté pour les modèles de Markov cachés comme techniques de base dites aussi statistiques et probabilistes. Ainsi, à partir du modèle des chaînes de Markov standard, nous avons proposé deux extensions qui concernent trois principaux volets et qui représentent nos principales contributions :

- i) l'intégration d'un raisonnement spatio-temporel dynamique permettant de prendre en compte l'évolution temporelle des scénarios. Le temps étant un paramètre important pour discriminer différents types de scénarios [AME14a].
- ii) l'intégration d'un raisonnement en arborescence hiérarchique permettant de prendre en compte la structure interne et organisationnelle des événements. Cette structure est supposée hiérarchique dans notre modèle.
- iii) la gestion d'incertitude, d'imprécision et d'incomplétude des données dans un cadre crédibiliste à travers l'architecture de tout le système proposé [AME13].

S'agissant de la gestion d'incertitude, il convient de noter deux principaux niveaux de gestion d'incertitude du système, qui s'avèrent très importants face à la sensibilité du domaine médical. Le premier niveau concerne les données médicales incertaines et imprécises et des connaissances médicales incomplètes [SAS04]. Le second niveau, quant à lui, traite de la gestion de l'imprécision des capteurs de vision de l'incertitude des processus de vision. Un apport significatif des techniques proposées est leur genericité qui leur confère une robustesse caractérisée par leur capacité d'adaptation à d'autres problèmes avec des applications diverses de reconnaissance de séquences d'états [Ram07]. Le domaine d'application qu'est celui médical est tant bien une nouveauté qu'une originalité. Il ouvre des voies d'investigation en terme de

recherche vers la mise en place d'une boîte noire médicale qui constitue une solution très prometteuse pour améliorer les systèmes de surveillance actuel dans les unités de soins intensifs médicaux en termes de qualité de service.

3. Organisation de la thèse

Le manuscrit de thèse est structuré en cinq chapitres :

Chapitre 1 - Les systèmes de vidéosurveillance intelligents ont été présenté à travers leurs enjeux, leurs fonctionnalités et surtout leurs diverses applications. A l'issue de cette présentation, un bilan est dressé et montre l'intérêt et l'utilité de ces systèmes pour le monitoring d'activités humaines en milieu médical. Ensuite, le contexte de la surveillance en milieu médical que constitue l'application visée par nos travaux de recherche est présenté. Le système de vidéo surveillance proposé pour suppléer l'existant en termes de surveillance en unité de soins intensifs est décrit. Enfin, nous avons justifié l'opportunité d'une investigation vers la mise en place d'une boîte noire médicale en milieu hospitalier.

Chapitre 2 - Ce chapitre est un état de l'art et a pour but de présenter les différentes étapes d'un système de vidéo surveillance intelligente en partant des traitements bas niveau aux traitements haut niveau. Cependant, les traitements haut niveau pour la reconnaissance de scénario dans des séquences vidéo ont été bien développés. Nous avons groupé en deux catégories ces traitements qui englobent les méthodes orientées symboliques et les méthodes orientées numériques. Ensuite, à partir d'un bilan sur les techniques bas niveaux, nous avons effectué un choix de notre approche en nous basant sur les aspects temps-réel et surtout la capacité d'adaptation de la technique choisie aux outils que nous exploitons dans la couche haute d'interprétation vidéo. Enfin, les outils de gestion d'incertitude ont été aussi présentés dans ce chapitre.

Chapitre 3 - Les modèles de Markov cachés et leurs principales extensions ont été étudiés. Dans un premier temps, une description des différentes topologies de ces modèles et leur exploitation pour la reconnaissance de scénarios dans des séquences vidéo est présentée. Dans un second temps, une extension des modèles de Markov cachés qui intègre une couche de gestion de la structure hiérarchique et de l'évolution

temporelle des événements (scénarios) a été proposée en se basant sur les chaînes de Markov cachés hiérarchiques et le modèle semi-markovien. Ainsi, l'intégration d'un paramètre probabiliste de gestion explicite du temps passé par le modèle dans un état est exploitée dans le modèle proposé et a été présentée. Enfin, après avoir décrit, l'approche proposée qui est basée sur les modèles de Markov cachés hiérarchiques et temporel, son exploitation pour la reconnaissance de scénario dans les vidéos médicales est justifiée et exposée. Un algorithme avec une architecture en deux couches pour la gestion de la structure hiérarchique et l'évolution spatio-temporelle des scénarios a été proposé pour la reconnaissance de scénario [AME⁺14].

Chapitre 4 - Il fournit une vue globale des outils de gestion d'incertitude et la justification de notre choix porté sur un outil de gestion d'incertitude de type crédibiliste. Ensuite, le Modèle de Croyances Transférables (MCT) est présenté et nous décrivons par ailleurs les opérations de base pour la fusion d'informations dans le cadre des modèles de croyances transférables. Les modèles de Markov crédibilistes ont été décrits puis les mécanismes algorithmiques sous-jacents aux modèles de Markov probabilistes sont en partie généralisés aux fonctions de croyance. Une gestion du temps de séjour du modèle dans un état et une gestion de la hiérarchie structurelle des états sont aussi intégrées dans les modèles de Markov cachés crédibilistes présentés. A cette fin, un algorithme est proposé pour filtrer des fonctions de croyances évoluant au cours du temps ayant une structure hiérarchique pour reconnaître des événements importants dans des séquences vidéo.

Chapitre 5 - Ce chapitre est essentiellement expérimental. Il commence par décrire le système expérimental et la base de données vidéo médicales exploités pour tester nos algorithmes de reconnaissance dans des séquences vidéo. L'approche de prétraitement des images, proposée est présentée et les résultats de son exploitation sont aussi présentés et analysés. Les approches développées dans les chapitres 3 et 4 sont exploitées et évaluées sur une base de vidéos médicales et comparées aux modèles de Markov cachés probabilistes.

Chapitre 1

Vers un système de vidéo-surveillance multicaméras intelligent en Unités de Soins Intensifs

Introduction

La vidéosurveillance intelligente est l'un des domaines de recherche les plus actifs de la vision par ordinateur cette dernière décennie. Elle offre une grande variété d'applications à la fois dans des environnements publics et privés, tels que la sécurité intérieure, la prévention du crime [WTK⁺11, AJC11], le contrôle du trafic [SCW10, YZY⁺11], la prédiction d'accident [JGH⁺10, LM11], le monitoring des activités sportives, le monitoring d'activités humaines pour le maintien à domicile des personnes âgées [RMSA⁺11, ZBT10, RCC⁺13] et la surveillance des patients en milieu médical [AME14a]. Nous présentons, d'une part, dans ce chapitre, les principales applications avec un bilan des caractéristiques et des spécificités de chaque type d'application afin de montrer l'intérêt de la vidéosurveillance intelligente en environnement médical. D'autre part, nous présentons les Unités de Soins Intensifs (USI) à travers leurs structures organisationnelles, leurs caractéristiques, et les spécificités des données manipulées dans ces unités tout en mettant en exergue la particularité de la surveillance en USI par rapport aux autres types de surveillance [SAS04]. Ensuite, nous faisons une analyse des systèmes de surveillances actuels en USIs dans les secteurs médicaux tels que la cardiologie et l'anesthésie-réanimation afin de mettre en évidence les limites de ces derniers [SAS04, FRP11, SGM07]. Enfin, nous décrivons l'architecture du système de surveillance visuelle dans un contexte multicaméras que nous proposons pour compléter le système de surveillance sonore en USIs tout en introduisant le concept de la Boîte Noire Médicale (BNM) comme une nouvelle composante de ces systèmes [AME⁺14].

1.1. Les systèmes de vidéosurveillance intelligents

Les systèmes de vidéosurveillance intelligents ont pour principal objectif d'extraire efficacement et en temps-réel des informations utiles à partir d'une quantité importante (volumineuse) de vidéos recueillies par les caméras de surveillance qui peuvent être en réseau. En effet, l'expansion des systèmes de surveillance vidéo a entraîné l'émergence de nombreuses applications dans divers domaines importants de la technologie tels que la sécurité, des lieux publics intérieurs (parcs de stationnement, centre commercial, aéroport, gare) et des lieux publics extérieurs (trafic routier et ferroviaire) [BXL11, LJK11] dans le cadre de la prévention et de la gestion du crime [RMSA⁺11]. La surveillance du trafic pour la détection et la prévention d'accident [AJC11], l'analyse et le contrôle de la circulation [SCW10, YZY⁺11], la gestion de la circulation aériennes [LWJ11], le contrôle du trafic maritime [MLC⁺10, SLT11], des statistiques démographiques, qui demandent de compter les objets mobiles (piétons, véhicules) [CFP⁺10] à travers l'identification, le suivi et la classification d'objets (véhicules par exemple et tout objet d'intérêt). Récemment, dans le cadre des applications de télésurveillance en milieu médical, l'assistance des personnes âgées dans les maisons de retraite [JGH⁺10, LM11] est en plein essor. Tout ceci est rendu possible par l'exploitation de techniques robustes de détection de mouvement, d'analyse et d'interprétation d'activité ou de la détection de comportement humain [TFS⁺10, ALD11]. Face à cette large gamme d'applications et en raison de la disponibilité croissante de capteurs à bas prix tels que les caméras intelligentes et embarquées, un besoin de plus en plus croissant pour la sécurité, la vidéosurveillance témoigne d'un intérêt particulier qui n'est plus à démontrer. Par exemple, aujourd'hui, nous avons des dizaines de milliers de caméras dans une ville (1000 caméras à Paris depuis 2010 [Luv10, Hei11]) qui collectent une quantité énorme de données sur une base quotidienne et qui doivent être analysées et interprétées par le système et validées au besoin par un opérateur humain pour la phase de prise de décision. Cette section passe en revue l'évolution des technologies récentes et pertinentes à travers quelques applications de la vidéosurveillance intelligente. Pour mieux cerner l'attrait des systèmes de vidéosurveillance intelligents [CFA13, VVA05, DRH11], nous introduisons

dans la section suivante les différentes fonctionnalités technologiques implémentées par un tel système afin de dégager les applications qui en découlent.

1.1.1. Les principales fonctionnalités en vision par ordinateur

Les principales fonctionnalités technologiques d'un système de vidéosurveillance intelligent dans un contexte multicaméras peuvent se résumer notamment au calibrage des caméras, la détection automatique et le suivi d'objets mobiles dans des séquences vidéos, la reconnaissance et la classification d'objets d'intérêt, l'analyse d'activités humaines et l'interprétation vidéo pour la compréhension de scène. La Figure 1.1 présente quelques principales fonctions technologiques proposées en vision par ordinateur par les systèmes de vidéosurveillance intelligents multi-caméras tout en mettant en exergue les interactions qui peuvent exister entre ces différents modules technologiques [Wan12].

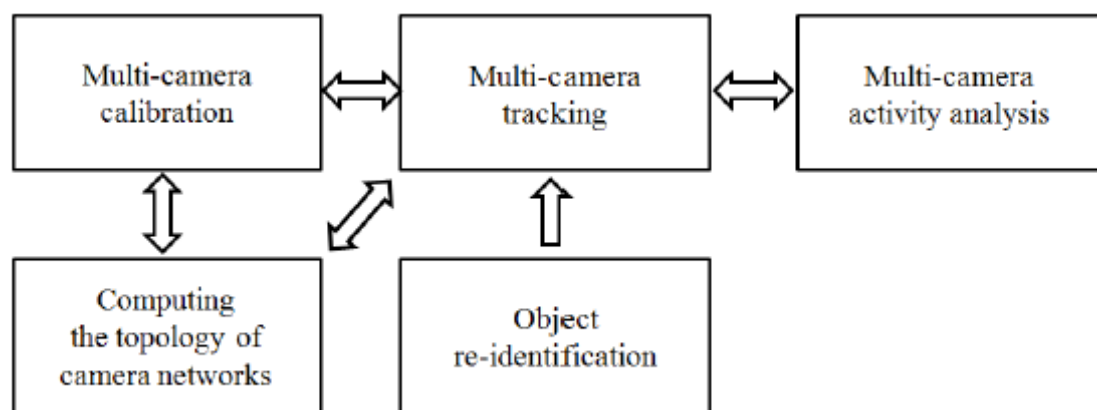


Figure 1.1. Illustration des principales technologies de vidéosurveillance multicaméras. Les flèches indiquent la circulation du flux d'information entre les différents modules [Wan12].

Le calibrage d'un système multi-caméras avec des différentes vues des caméras dans un repère global et unique de coordonnées images. Dans de nombreux systèmes de surveillance, le calibrage représente une phase de prétraitement très importante pour l'efficacité des résultats des autres couches d'analyses multi-caméras supérieures.

La topologie du système de vision à perception multicaméras (l'architecture du réseau de caméra) qui permet d'identifier si les champs de vues des caméras se chevauchent dans l'espace adjacent et décrire le temps de transition entre des objets des différentes vues de la caméra.

La ré-identification d'objet qui consiste à faire correspondre deux zones d'images observées dans différents champs de vues de la caméra et de reconnaître qu'ils appartiennent au même objet ou pas, en n'utilisant que des informations d'apparence d'objets sans raisonnement spatio-temporel.

La détection et le suivi multicaméras qui consistent à détecter des mouvements d'objets mobiles de la scène et de les suivre à travers les différents champs de vues du réseau de caméra.

L'analyse et l'interprétation de l'activité humaine ou de tout objet d'intérêt dans une scène vidéo via le système multicaméras. Elle consiste à reconnaître automatiquement différentes catégories d'activités et de détecter des événements d'intérêts en combinant ou en fusionnant (appel aux outils et techniques de combinaison ou de fusion d'information) les informations issues de plusieurs caméras (sources d'informations).

Ces différents modules technologiques [AC09, VVA04] peuvent être classés en deux niveaux de traitement à savoir : le module de traitement bas niveau qui concerne le calibrage, l'extraction de caractéristiques, la détection et le suivi d'objets mobiles et le module de traitement haut niveau qui est relatif à la reconnaissance, l'analyse et l'interprétation vidéo. Ces deux couches bas et haut niveaux forment l'ensemble des blocs fonctionnels nécessaires à tout système de surveillance intelligent multicaméras et distribués au besoin. Nous nous proposons dans la section suivante de présenter quelques principales applications de la vidéosurveillance intelligente afin de dresser plus tard un bilan des caractéristiques et des spécificités de chaque type d'application. Cette liste d'applications n'est pas exhaustive mais elle met en avant tout simplement la diversité de ces applications et l'importance de leurs domaines d'utilisation.

1.1.2. Une pluralité d'applications dans divers domaines

La vidéo-surveillance couvre un large champ d'applications. Grace aux récents progrès technologiques, elle répond aujourd'hui aux besoins d'un marché mondial en pleine évolution. À son origine, la vidéosurveillance a d'abord été utilisée par les services publics (police, transports, administrations, banque et commerce). Elle fut ensuite adoptée par des entreprises désirant protéger des actifs stratégiques, telles que les raffineries, les centrales nucléaires, les barrages fluviaux, les industries agroalimentaires et les complexes pharmaceutiques [VVA04]. Les applications des systèmes de vidéosurveillance concernent de nombreux domaines avec une liste non exhaustive des plus connues pour laquelle nous faisons une présentation globale. La vidéo-surveillance intelligente offre de nombreuses possibilités en matière de sécurité des biens et des personnes. Le traitement d'images automatisé et la vidéo analytique en temps réel, permettent par exemple de «*tracker*» des individus suspects et/ou de détecter des incidents et des événements importants (comportements anormaux) d'une scène. Nous présentons trois types d'applications auxquels les enjeux de la vidéosurveillance sont liés :

- la «*sécurité*», un domaine dans lequel les recherches en système de vidéo surveillance intelligente ont explosé depuis par exemple les événements de 2001 aux USA ;
- le «*monitoring*» d'activités sportives par la reconnaissance d'activités sportives en fonction du type de sport ;
- le «*monitoring*» d'activités humaines en milieu médical (les patients, les infirmiers et les médecins) ou la télésurveillance des domiciles de personnes âgées.

1.1.2.1. Une solution de sécurisation des biens et des personnes

L'un des premiers domaines dans lequel les applications des systèmes de vidéosurveillance intelligente ont émergé est celui de la sécurité. Le développement des actes de criminalités a joué un rôle très important dans la croissance de la surveillance des lieux publics et privés. Pour cela, de nombreuses applications pour la sécurité des biens et des personnes ont vu le jour, parmi lesquelles nous avons la prévention et la

gestion du crime par la vidéosurveillance des lieux publics intérieurs comme les centres commerciaux et les parkings, et la vidéosurveillance extérieure comme la surveillance et le contrôle des infrastructures critiques et des autoroutes [VVA05]. Aujourd'hui, elle est devenue de plus en plus fréquente dans plusieurs types d'activités humaines comme par exemple les transports en commun, tels que les aéroports [WES95], les environnements maritimes et portuaires [PSR98, Avi03], les chemins de fer, les infrastructures souterraines [PSV03, Nwa98, RD00], et les autoroutes pour l'enquête sur le trafic [PT98, FMW00]. Les lieux publics tels que les banques, les supermarchés et les hypermarchés, les grands magasins [CGP+04] et les parkings de stationnements [GRJ02, MFS03], la surveillance en milieu industriel pour le contrôle de qualité dans de nombreux procédés industriels, source de preuve pour des applications médico-légales [GB00] et la surveillance à distance dans des applications militaires et des sites militaires, etc. représentent la cible de la vidéosurveillance. Par exemple, depuis les événements récents en termes d'insécurité dans le monde, y compris les attaques terroristes majeures, une demande accrue de sécurité dans la société s'enregistre. Cela oblige les gouvernements à faire de la sécurité et active une priorité dans leurs politiques. Cela se traduit par le déploiement de systèmes de vidéosurveillance à grande échelle un peu partout dans le monde. Par exemple, c'est le cas du métro de Londres et l'aéroport d'Heathrow qui ont plus de 5000 caméras chacun [HHD00]. Pour cela, un système de vidéosurveillance intelligent dans le but de surveiller les voleurs potentiels dans les grands magasins, a été développé et se compose d'un réseau de nœuds de capteur de vision. Chaque nœud du réseau est composé d'une caméra de type PTZ connectée à un ordinateur et une console centrale pour être utilisé par un opérateur humain. Le système signale la présence d'un individu marqué à l'intérieur du bâtiment en supposant que le trafic d'êtres humains est rare (une hypothèse qui devient moins valide pour une foule).

1.1.2.2. Une solution de monitoring d'activités sportives

Une seconde catégorie d'applications de la vidéosurveillance intelligente concerne la surveillance à distance des activités humaines telles que la participation à des matches de football [XLO04] ou d'autres activités humaines [KHM+00, HHD00] que ce soit dans

le domaine sportif ou pas (surveillance des places assises d'un moyen de transport, surveillance de comportement d'une foule). Nous aborderons principalement le monitoring des activités sportives telles que le football [XLO04, DS04], le basket-ball, le tennis [Kij03] et le base-ball [CLL07]. Les vidéos de sports représentent une proportion non négligeable des bases de vidéos et les enjeux liés à l'analyse automatique de ces vidéos sont cruciaux et importants pour les professionnels, par exemple pour l'annotation [BBN04, MCJ+05, RCKL07]. La plupart des travaux de recherche se sont focalisés sur l'analyse des vidéos de sports diffusées à la télévision. La première raison est que celles-ci véhiculent différents types d'informations pertinentes pour leur analyse comme les paroles des commentateurs, les indications textuelles données par les chaînes et les données visuelles représentées par les images. La seconde raison est qu'il existe de nombreuses applications autres que la gestion des bases de vidéos comme l'analyse de performances et des tactiques [She06] afin de donner des statistiques aux téléspectateurs, joueurs et entraîneurs [YHYC05, BJA07], la réalité augmentée [KKO03, YYHL04, She06] afin de créer des environnements 3D pour visualiser des gestes techniques ou des fautes et le suivi d'objets comme les balles de tennis ou les ballons de football [YLXT06, YLL+03, BJA07]. Cependant, une des applications majeures reste, sans nul doute la création de résumé [Ram07].

En 2007, Lien et al. ont proposé un système de vidéosurveillance intelligent basé sur un système de reconnaissance qui exploite les modèles de Markov cachés dans une vidéo de sport de baseball pour définir et détecter les frappes réussies. Pour cela, quatre événements prédéterminés ont été définis et sont à détecter. Leur approche consiste à extraire une image clé pour chacun des plans obtenus par la segmentation de la vidéo. Ensuite, une estimation de mouvement global dans la vidéo est effectuée. La taille et la position des objets sur les images clés sont calculées et, enfin, une classification des plans dans l'une des huit classes est effectuée. Le modèle de Markov caché de quatre états permet de modéliser les scénarios et vise à reconnaître l'un des quatre événements prédéterminés.

D'autres travaux de recherche similaires ont été réalisés dans [CLL07] pour la reconnaissance d'actions dans des vidéos de football [DS04] et du match de tennis [Kij03]. De même, les travaux de recherche de E. Ramasso sur la reconnaissance de

séquence d'état dynamique [Ram07] dans le cadre d'analyse de mouvements humains [MHK06] pour la gestion de base de vidéos sportives et des interfaces Homme-Machine [JS05] ont permis de développer une solution de vidéo surveillance intelligente pour reconnaître des actions et des activités humaines dans les vidéos d'athlétisme. En effet, l'application permet de reconnaître les sauts d'athlétisme dans des séquences vidéo. Ce type de système pourrait par exemple être utilisé pour trier les événements sportifs sur un enregistrement d'émissions télévisées pour les journalistes, ce qui pourrait faciliter une opération de montage vidéo. Ce type d'application est aussi intéressant pour produire des résumés de scènes sportives dans le cadre de la réalisation de magazines de sports télévisés.

En somme, plusieurs applications de monitoring d'activités sportives ont été proposées parmi lesquelles nous avons : le football européen [ABC+03, SOMM04, BLZ+07, WCX+07], le football américain [LEPS04, DF07], le cricket [KS06], le volleyball [XMZY05], les jeux de raquettes [ZHX+07] comme le tennis de table [ZCDZ07] ou le tennis sur court [KGG+03, LMI06], le baseball [LWH03, LEPS04, CH06, LCL07], le basketball [ZXR+07], l'aérobic [BD01], le badmington [CZ07], le sumo [LEPS04], le snooker [DRK03], le patinage et l'athlétisme [Pir05, PBY06], la boxe [BCJ05, LMI06], la course de chevaux [OND+05, TCP04], la danse [EBMM03], le rugby [TCP04, BO05], le golf, le ski et le bowling [LWH03], le hockey [LL06], la natation [TDX+06], le plongeon, la gymnastique, le judo et le cyclisme [MCJ+05]. Un état de l'art assez complet a été proposé dans la thèse d'E. Ramasso [Ram07].

1.1.2.3. Une solution de monitoring d'activités humaines en milieu médical

Une troisième catégorie d'application de la vidéosurveillance intelligente concerne le monitoring d'activités humaines en milieu médical (par exemple la télésurveillance pour le maintien à domicile des personnes âgées [ZBT10]). Généralement et pour des besoins primordiaux de sécurité, les centres médicaux sont dotés de système de vidéosurveillance mais adapté à leur contexte. Ainsi, dans le but de sécuriser les espaces et l'accès aux bâtiments d'une clinique, d'un hôpital, d'un cabinet médical et autres établissements de santé, un recours à la vidéosurveillance est fait. Une approche consiste à installer un réseau de caméras dans les différents services (urgences,

radiologie, urologie, anesthésie-réanimation maternité et autres) de l'hôpital pour surveiller les usagers. Par exemple, pour les maternités, des possibilités de solution de vidéosurveillance permettent aux parents de voir leur nouveau-né à distance. Les parents peuvent ainsi garder un contact visuel avec leur enfant. La vidéosurveillance dans le secteur médical offre de nombreux avantages à savoir :

- les systèmes de vidéosurveillance peuvent permettre de sécuriser l'établissement médical (contrôler les accès aux locaux spécifiques),
- le système de vidéosurveillance, une fois installé dans les pouponnières, permet aux parents du nouveau né de garder un lien avec leur bébé,
- la vidéosurveillance permet d'assister les interventions chirurgicales ou autres actes cliniques médicaux.
- Enfin, le système permet à partir d'une analyse à postériori de la base de données vidéos d'amélioration des méthodes de travail, meilleure gestion des ressources humaines voir matériels. Il participe à l'hygiène l'ergonomie et l'organisation de la circulation de l'USI.

Ces solutions de vidéosurveillances en milieux hospitaliers sont très utilisées depuis une dizaine d'années dans les pays tels que la Grande Bretagne, l'Angleterre, le Canada et les Etats Unis. En France, et plus précisément à Sophia-Antipolis, le projet *GERHOME* (pour *GERontology at HOME*) développé depuis plusieurs années par le Centre scientifique et technique du bâtiment (Cstb) est une application majeure de la vidéosurveillance intelligente en environnement médical. L'un des objectifs du projet *GERHOME* est de développer une solution visant à fournir un monitoring fiable d'une personne âgée vivant seule et pour le maintien à domicile des personnes âgées [RCC+13]. Il s'agit d'améliorer le suivi médical et pouvoir informer sans délai de l'inactivité dans l'habitat, lancer une alerte en temps réel, à la demande de la personne âgée, ou encore de façon automatique (en cas de chute ou de malaise, par exemple). La mise en place de ces solutions doit cependant faire face à différentes contraintes telles que le respect de la vie privée, la facilité d'installation, de configuration et de maintien et le fait de fournir des informations et des services adaptés aux utilisateurs. Une attention particulière est donnée aux aspects de confidentialité et de protection de la vie privée.

Le projet GERHOME utilise des technologies «domotiques» intelligentes qui visent à assurer l'autonomie, le confort de vie, la sécurité, la surveillance et l'assistance à domicile. L'expérimentation et l'évaluation de ces services numériques vont permettre aux personnes âgées de continuer à demeurer dans leur cadre habituel, dans des conditions de sécurité optimales. De même, dans un contexte d'application médicale similaire, il nous semble opportun d'intégrer ces systèmes de vidéosurveillance intelligents aux systèmes de surveillance existant dans les unités de soins et surveillance des patients dans certains secteurs médicaux tels que la cardiologie et la réanimation, etc. Cette intégration permettra d'avoir des alarmes visuelles plus informatives que celles sonores de l'existant afin d'améliorer la qualité de la surveillance [AME14a].

Les défis et les fonctionnalités envisagés pour un tel système dans ce contexte d'application le rendent plus complexe. Afin de mettre en exergue les spécificités de la vidéo surveillance en milieu médical et spécifiquement en unités de soins intensifs, un bilan comparatif des spécificités liées à chaque domaine d'application de la vidéosurveillance intelligente, que nous avons abordé dans ce document est fait dans la section suivante.

1.1.3. Bilan des spécificités liées à chaque type d'application

Après avoir fait un succinct panorama des applications de la vidéosurveillance intelligente dans quelques domaines prépondérants, nous nous proposons de faire un bilan du point de vue des spécificités de chaque type d'application afin de mettre en exergue les particularités de la vidéo surveillance en milieu médical mise à part le cas que constitue la surveillance en unités de soins intensifs [SAS04].

Généralement, un système de vidéosurveillance est intimement lié à la nature ou à un type de scène (simple ou complexe) et par conséquent au contexte de l'application. Dans le cadre d'une analyse de scène, les caractéristiques des données exploitées et les connaissances à priori sur les événements à détecter par le système s'avèrent très importantes pour l'efficacité du système. Nous tenterons de comparer ces applications par rapport à trois critères généraux que sont les contraintes applicatives, les caractéristiques des données manipulées et les connaissances a priori disponibles pour

construire et modéliser le système et les événements de la scène à reconnaître (voir tableau 1.1). Parmi les critères liés aux contraintes applicatives, nous avons :

- des contraintes techniques, comme par exemple, la calibration de caméra ou de réseaux de caméras, les éventuelles distorsions provoquées par les capteurs, des taux de rafraîchissement faibles dans certaines situations, etc.
- des contraintes environnementales comme le manque de résolution dans certaines situations, les changements d'illumination, etc. Ce sont des problèmes récurrents en vision par ordinateur et il n'existe pas encore de solutions «générique». Ils se traitent en général au cas par cas selon la scène observée, ce qui rend la généralisation d'un système de vidéosurveillance d'autant plus difficile.
- des contraintes sémantiques : les systèmes mis en œuvre sont généralement ciblés pour un type ou une plage d'évènements bien spécifiques. Par exemple, la détection de personnes trop proches du bord d'un quai ou la détection de présence de personnes sur la chaussée, etc.

S'agissant de la nature des données traitées par le système, elle est relative à la quantité des données (nombre de variables et de mesure), et leur qualité (non bruité, moyennement bruité et trop bruité). De même pour les connaissances a priori, elle concerne le fait que ces connaissances soient disponibles ou non. Dans les cas où elles sont disponibles, en quelle proportion elles le sont (faible, moyenne ou élevée)? Sont-elles formalisables à des fins d'apprentissage automatique? Au regard des résultats que montre le Tableau 1.1, nous pouvons constater que pour tous les domaines d'applications, les contraintes applicatives que ce soit de types technique, environnemental et sémantique sont présentes et élevées mise à part le cas des applications dans le domaine médical pour lesquelles elles sont très élevées.

En termes de caractéristiques et de la nature des données, nous avons des données bruitées dont les causes sont multiples et la quantité peut être élevée et parfois très élevée pour toutes les applications. Pour les connaissances a priori, elles peuvent être disponibles, mais une complexité de formalisation assez importante existe pour toutes les applications. De façon plus spécifique, pour les applications du domaine médical, les connaissances a priori existent mais en proportion moyenne et sont incertaines et incomplètes et par conséquent leur formalisation s'avère très ardue.

Types et domaines d'applications	Contraintes applicatives			Caractéristiques des données traitées		Connaissance a priori	
	Techniques	Environnementales	Sémantiques	Quantité	Qualité	Proportion de disponibilité	Possibilité de formalisation
La sécurité des biens et des personnes	Oui	Elevé	élevé	Elevé	Bruitées	Elevé	Oui mais parfois complexe
Le monitoring d'activités sportives	Oui	Elevé	élevé	Elevé	Bruitées	Elevé	Oui mais parfois complexe
Le monitoring d'activités humaines en milieu médical	Oui	Elevé	très élevé	Elevé	bruitées et imprécises	moyenne, incertaine et incomplète	Oui, mais très ardue car trop complexe

Tableau 1.1. Tableau comparatif des caractéristiques et spécificités de quelques applications de vidéosurveillance intelligente selon le domaine d'application.

1.2. Présentation d'une Unité de Soins Intensifs

Une unité de soins intensifs est une section spécialisée des services médicaux tels que l'anesthésie-réanimation, la cardiologie et les urgences où l'on retrouve des patients gravement malades qui reçoivent des soins particuliers continus et sont surveillés en temps réel par une équipe de médecins avec une équipe d'infirmiers via un système d'assistance et d'aide à la décision composé d'équipements médicaux. En raison de l'état de santé très critique des patients, ces unités sont généralement dotées de plusieurs équipements pour une assistance respiratoire aux patients [SAS04, SDG05]. Par exemple, aux unités de soins intensifs de la section cardiologique qui est le contexte expérimental, on trouve généralement des patients qui ont des problèmes d'insuffisance respiratoire causée par une détérioration des propriétés mécaniques du poumon, le système nerveux central et surtout le système cardio-vasculaire. Ainsi, un mauvais fonctionnement du système respiratoire humain est lié à une défaillance des organes tels que les muscles respiratoires, les muscles neurologiques, le ventricule gauche ou droit et d'autres comme des anomalies au niveau des échanges gazeux et des facteurs psychologiques. Le dysfonctionnement de ces derniers organes chargés d'assurer ces fonctions est la cause fondamentale de l'insuffisance respiratoire [FRP11]. Par conséquent, pour aider le patient à se maintenir en vie, une assistance respiratoire partielle ou totale est obligatoire en fonction de la gravité de l'état dans lequel se trouve le patient. Pour ce faire, l'assistance respiratoire justifie d'une surveillance importante et très particulière.

Dans la suite de cette section, nous présentons la structure organisationnelle, les composantes d'une unité de soins intensifs. Nous mettons aussi en exergue l'incomplétude des connaissances médicales à partir des caractéristiques et les spécificités des informations et des données manipulées en unité de soins intensifs.

1.2.1. Structure organisationnelle et composantes d'une USI

La médecine intensive comprend le diagnostic, la prévention, les soins et le traitement de toutes les formes de défaillances aiguës des fonctions vitales chez le patient en danger. Cette médecine est pratiquée par une équipe formée et spécialisée se

composant de personnel médical, infirmier et paramédical [SAS04]. Elle est exercée dans un emplacement défini et approprié à cet effet avec l'aide d'équipements technologiques médicaux de pointe en pleine rénovation.

Dans les unités de soins intensifs des services médicaux comme la cardiologie, séjournent des patients ayant de sévères troubles sanitaires et qui reçoivent de ce fait, des soins spécifiques et sont surveillés en temps-réel à cause de leur instabilité sanitaire par un staff médical. Cette équipe médicale multidisciplinaire est composée parmi tant d'autres spécialistes, d'intensivistes (médecins réanimateurs ou spécialisés dans les soins de maladie grave), infirmiers, anesthésistes, chirurgiens, nutritionnistes, qui offrent des soins avec l'aide des équipements technologiques médicaux déployés aux soins intensifs [FRP11].

D'une part, en raison de la gravité de l'état de santé des patients qui est la plupart du temps liée à une insuffisance respiratoire, l'assistance et la surveillance de ses patients, les salles de soins intensifs sont généralement équipées de ventilateurs mécaniques pour aider à respirer, moniteurs cardiaques, des stimulateurs cardiaques externes et les défibrillateurs, matériel de dialyse pour des problèmes rénaux, les drains et un large éventail de médicaments [SAS04, FRP11, SDG05].

D'autre part, en raison de l'absence d'un garde-patient et du nombre limité du personnel médical chargé de surveiller les patients qui peuvent se trouver dans des salles de soins séparées, les salles de soins intensifs sont intégrées d'un système de surveillance constante en temps-réel des fonctions vitales des patients, et des fonctions corporelles. Afin d'identifier les problèmes qui peuvent survenir et nécessiter une attention ou un traitement d'urgence, les médecins procèdent à des évaluations périodiques de l'état cardiaque du patient, le rythme respiratoire, la production d'urine et les niveaux de sang [FR11, SGM07].

Dans certains cas, les patients peuvent avoir besoin d'exigences particulières pour le suivi. C'est le cas des patients qui sont admis à l'unité de soins intensifs pour observation après la chirurgie. Ces patients peuvent avoir des cathéters placés pour détecter les changements hémodynamiques (modifications de la pression artérielle) ou exiger une intubation endo-trachéale pour les aider à respirer, avec le tube respiratoire

relié à un ventilateur mécanique. Afin d'effectuer un traitement correct, la disponibilité d'informations détaillées est d'une importance vitale.

Les médecins non seulement analysent les données affichées par les équipements de surveillance, mais aussi analysent les résultats des examens de laboratoire et de radiologie. Ils évaluent d'autre part, à partir de l'historique des états cliniques du patient et les dosages des médicaments à administrer à ce dernier, tout en tenant aussi compte d'informations spécifiques du patient telles que son âge, son poids et sa taille. Enfin, les médecins analysent aussi les avertissements d'interactions médicamenteuses afin de poser des diagnostics [MFB02].

Pour une meilleure mise en œuvre du plan de traitement médical pour un meilleur résultat de la prise en charge du patient et éviter au maximum les erreurs, les informations mentionnées ci-dessus sont extrêmement importantes pour les médecins. Ce système de surveillance, constitue de ce fait, une aide à la décision pour les médecins surveillants dans leur tâche de surveillance. Une salle de soins intensifs bénéficie des équipements ci-dessous et organiser typiquement comme à la Figure [SAS04]:

- un ventilateur pour assister ;
- un moniteur pour la surveillance des paramètres (visualisation des signaux, réglages de seuils) ;
- d'autres moniteurs (scanographe, moniteur cardio-vasculaire,) ;
- un poste central de recueil et de centralisation d'informations.

En raison des avancées technologiques en termes de construction d'équipements médicalisés ces dernières années, plusieurs systèmes et appareils médicaux ont été mis en place et ont permis d'améliorer la surveillance en salles de soins intensifs grâce à l'introduction d'une intelligence et de l'automatisme dans ces systèmes. A titre d'exemple de cette avancée technologique, nous avons les lits des patients plus industrialisés. Deux fonctionnalités principales caractérisent cette industrialisation du lit des patients à savoir [SAS04] :

- une fonction d'information en temps-réel sur l'état physiologique ou physique du patient ;

- une fonction d'alarme ou d'alerte prévenant de modification des paramètres surveillés.

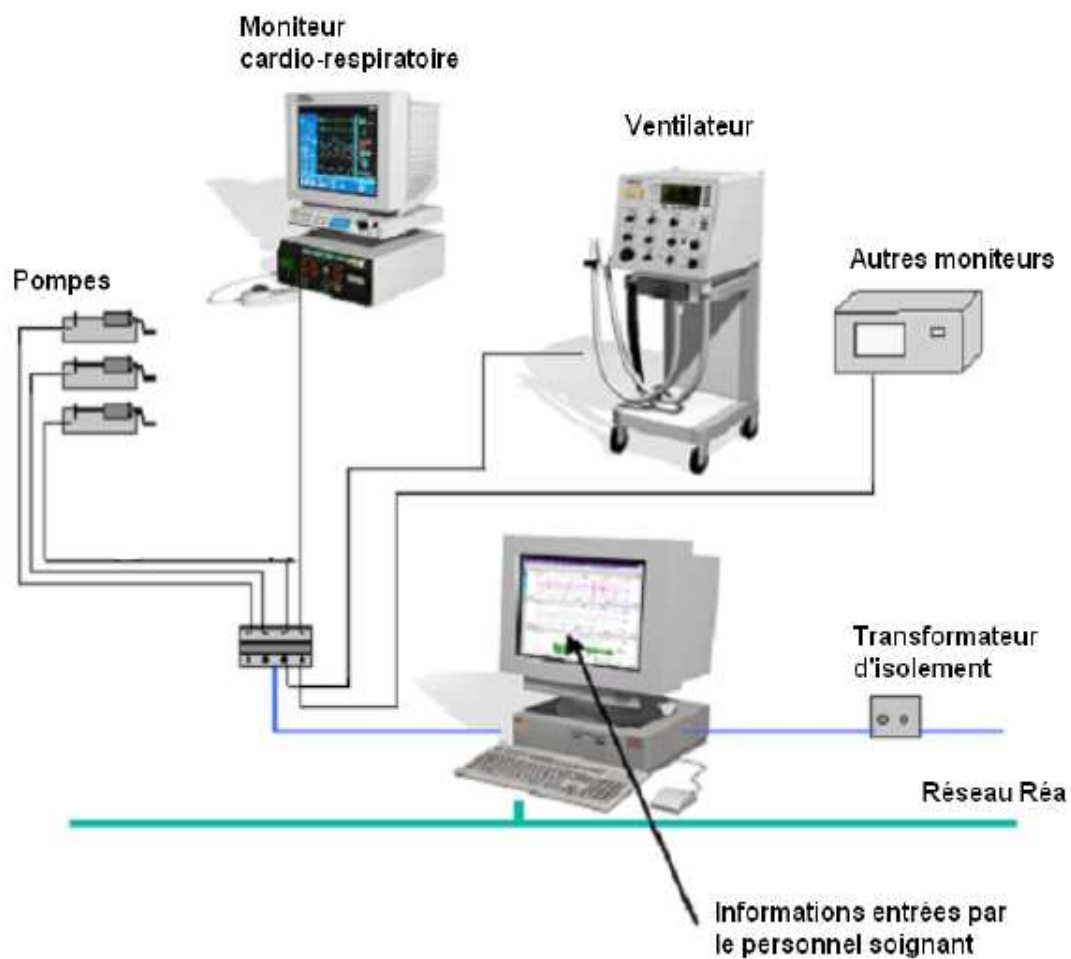


Figure 1.2. Illustration de l'organisation type d'une salle de soins intensifs [SAS04].

1.2.2. Caractéristiques et spécificités des données et des connaissances d'une USI médicale

La surveillance en milieu hospitalier et en particulier en salle de soins intensifs revêt un caractère spécifique et diffère des autres formes de surveillance dans d'autres domaines. Afin de mettre en exergue les spécificités de la surveillance en USIs, nous étudierons ce type de surveillance en faisant ressortir les caractéristiques et les spécificités tant bien des connaissances médicales que des données manipulées en unités de soins intensifs.

1.2.2.1. La surveillance des patients en unités de soins intensifs

Le terme Unité de Soins Intensifs (USIs) regroupe les unités d'anesthésie, de réanimation, les unités spécialisées de surveillance continue et les urgences médicales où une surveillance permanente et efficace des patients est importante pour maintenir les fonctions vitales du patient. Ces unités peuvent être parfois spécialisées pour un type de patient ou non. Par exemple, l'USI en cardiologie qui constitue le cadre expérimental de nos travaux est bien illustratif dans ce cas d'espèce. Chacune de ces unités est régie par des contraintes et des recommandations spécifiques. L'efficacité des traitements médicaux intensifs des patients qui sont dans un état instable et suffisamment grave nécessite une surveillance constante et continue [SAS04].

En effet, de nombreux patients séjournant en unités de soins intensifs sont surveillés et soignés en temps réel par une équipe médicale intensiviste. De même, plusieurs équipements médicaux permettent de surveiller et d'assister les patients. Chaque patient, que ce soit à son admission ou à sa sortie de l'USI, fait l'objet d'un rapport médical et en cas de transfert interne, le patient est accompagné de l'intégralité des documents qui le concernent (dossier médical, radiographies, etc.).

Les médecins de l'USI établissent un dossier médical pour chaque patient. En cas de transfert de patients de l'USI dans un autre hôpital, un rapport actualisé ainsi que des copies de tous les résultats et documents importants sont transmis avec le patient. Une liste de garde et de personnel permettent d'alerter rapidement les différents intervenants. Chaque appareil est pourvu d'une fiche de maintenance sur laquelle sont notifiées les dates de réparation et de contrôle technique.

Le minimum requis pour une salle de soins intensifs est de six lits et peut aller jusqu'à douze. Au-delà de douze lits, il est recommandé de subdiviser l'USI en blocs fonctionnels distincts. L'accès aux lits des patients pour le personnel soignant est distinct de celui réservé aux visiteurs. Les horaires de visite sont réglementés, mais doivent être conçus de façon libérale et les visiteurs sont tenus de s'annoncer avant d'être autorisés à entrer dans l'USI. En conséquence, il y a une quasi-inexistence de garde-patients aux soins intensifs. Généralement, la salle de soins est séparée par une vitre de la salle de surveillance [MFB02].

La surveillance doit être assurée, avec contact visuel englobant l'ensemble des patients de l'USI et un monitoring centralisé si un membre du personnel soignant n'est pas en permanence dans la chambre du patient. Un poste central de surveillance est situé au carrefour des différents locaux de soins et permet de superviser l'ensemble des patients. Les chambres sont disposées autour du lieu central de surveillance, d'où elles sont facilement accessibles. Le contact visuel peut aussi être établi au moyen d'une solution technique (surveillance vidéo).

Le contexte applicatif des systèmes de surveillance qu'est celui des USI est de plus en plus privilégié et exigeant en termes de spécificités du système de surveillance. Plusieurs outils d'aide à la décision sont déjà implantés dans les unités de soins et les cliniciens en font usage et en tiennent compte dans l'exercice de leur fonction. Au sein de l'organisation même de ces unités, on y retrouve principalement un système de surveillance automatique d'assistance de plus en plus intelligent centralisé dans une zone sécurisée et d'accès strictement interdit à toutes personnes étrangères.

Les différents paramètres surveillés chez le patient tels que les signes vitaux et physiologiques sont redirigés vers des équipements du poste central où sont disponibles les informations de l'ensemble des patients de sorte que le personnel soignant puisse être rapidement informé de l'état de l'ensemble des patients depuis un même lieu. Parmi ces diverses informations vitales surveillées chez le patient, nous avons deux grandes catégories que sont les paramètres relatifs aux signaux physiologiques qui sont en général des signaux analogiques et les signaux bruts qui sont des données de types numériques. C'est le cas, des valeurs des paramètres physiologiques tels que le rythme cardiaque (Signaux d'un ECG). Pour chaque paramètre physiologique surveillé, il existe un système d'alarme qui a deux seuils fixés et prédéfinis avec deux niveaux : bas et haut [SAS04]. Lorsque l'un de ces seuils dépasse celui des paramètres surveillés, une alarme est enclenchée par le système. L'ensemble des alarmes enregistrées sur des appareils ou des annotations saisies directement dans le collecteur de données par le médecin en charge représente les données de type symboliques.

Les alarmes en salles de soins intensifs ont été conçues pour surveiller l'état physiologique et clinique d'un patient afin d'alerter le médecin surveillant lorsqu'il y a

un changement dans la physiologie du patient ou pour alerter le personnel d'un problème de l'appareil [GM09]. Les alarmes se déclenchent quand une variable physiologique franchit un seuil de consigne [DB01].

Dans l'état de l'art, le problème d'alarmes multiples dans les unités de soins intensifs a été reconnu et discuté par les professionnels de santé depuis le début des années 1990 [ME95]. Les alarmes en fonction de seuils existants compromettent la qualité et la sécurité des soins aux patients en raison du taux élevé de faux positifs excessifs associés. Ces derniers obligent et amènent parfois les infirmières à éteindre certains appareils surveillant les seuils fixés pour les alarmes déraisonnablement élevés ou faibles. Et les médecins deviennent en conséquence insensibles aux sons. En outre, les alarmes ne correspondent pas toujours à la criticité de l'état du patient. Ce qui entrave la capacité des infirmières de réagir rapidement avec l'intervention clinique appropriée [NGK⁺12].

1.2.2.2. Caractéristiques des données en unités de soins intensifs

Les unités de soins intensifs sont des services spécialisés du domaine médical et par conséquent héritent partiellement des caractéristiques spécifiques des données médicales.

En général, les données médicales relatives à l'état d'un patient sont assez difficiles à analyser et posent des problèmes de confidentialité et d'accès. Ces derniers sont des problèmes d'ordre éthique, économique et social. Plusieurs aspects liés à la nature et à l'exploitation de ces données doivent être pris en compte lors de la phase exploratoire de ces dernières. Le problème de la confidentialité peut être pris en compte par une étape d'anonymisation soit en codant ou en cryptant ces données. L'accès (la propriété des données) est géré via l'attribution de droits d'accès d'utilisateur via un contrôleur de domaine puisque la plupart des systèmes de gestion des données médicales existantes aujourd'hui centralisent l'ensemble des données médicales disponibles au niveau d'un serveur ou un poste central dont l'accès est réglementé par des autorisations [SAS04].

Caractéristiques et spécificités des données en USIs

Dans les USIs, nous avons des données que sont les données de types analogique et numérique d'une part (paramètres physiologiques surveillés) et d'autre part des données de type symbolique (les alarmes). L'ensemble de ces différents types de données obtenues en salle de soins intensifs ont en commun des caractéristiques spécifiques par rapport à deux aspects: quantitative et qualitative [SDG05].

S'agissant du premier aspect, nous avons l'hétérogénéité et l'explosion quantitative de ces données caractérisées par une évolution temporelle et fréquentielle très élevée et par conséquent n'est pas du tout aisé à formaliser. L'aspect qualitatif est relatif aux caractéristiques générales et spécifiques des données médicales citées plus haut dans la section précédente : l'incertitude, l'imprécision et l'incomplétude. Enfin, l'aspect d'ambiguïté relative aux données médicales ne doit pas être négligé.

L'hétérogénéité et l'explosion quantitative de données en USI

La surveillance d'un patient en salle de soins intensifs peut impliquer la surveillance en temps réel et par conséquent conduit à l'analyse et l'interprétation de plusieurs informations issues de sources disparates et indépendantes. Ces informations ont généralement une évolution exponentielle qui est liée par exemple à leur fréquence d'acquisition relativement très élevée. Parmi ces informations, nous avons les paramètres physiologiques, les données d'analyses biomédicales, les données pharmaceutiques, les alarmes etc.

La pluralité et l'hétérogénéité d'informations médicales justifient l'explosion de flux de données médicales [SGM07]. Ceci est d'autant plus remarquable et vrai dans le contexte de l'application de vidéosurveillance (données images/vidéos) que nous proposons pour coupler les systèmes de surveillance en USI existant en guise de son amélioration.

A long terme dans cette application, plusieurs types de données issues de différents capteurs (de contact : physique et de vision : caméra) de natures différentes seront analysées et interprétées. Ceci explique l'hétérogénéité de ces données. Une fusion d'informations multi-sources est nécessaire pour améliorer la prise de décision

dans le cadre de la surveillance en USIs. Il est bien connu dans la communauté de recherche que la diversité des applications multimédias et la croissance notable des données multimédias en termes de quantité fait objet de thématique de recherche d'actualité.

L'incertitude, l'imprécision et l'incomplétude des données en USI

Les notions d'incertitude, d'imprécision et d'incomplétude des données médicales mettent en exergue la confiance à accorder aux données médicales. En effet, l'incertitude est omniprésente lorsqu'il s'agit de modéliser ou du moins d'expliquer le comportement d'un patient.

Dans les unités de soins intensifs, on retrouve plusieurs équipements d'assistance respiratoire et de surveillance des patients qui traitent plusieurs types de données et fournissent des informations par rapport à l'ensemble des paramètres surveillés chez le patient. Les résultats des paramètres surveillés obtenus sont des valeurs numériques qui proviennent des signaux physiologiques comme la fréquence cardiaque (ECG : électrocardiogramme). L'imprécision est inhérente aux données d'USI et se caractérise par le bruit d'acquisition et les imperfections liées à la qualité du capteur de mesure du paramètre.

En conséquence, nous sommes en présence de mesures imprécises, incertaines et incomplètes. La prise en compte de l'incertitude du système lorsqu'il lance une alarme peut également être une information riche de sens pour que les cliniciens organisent au mieux leurs interventions, mais également pour qu'ils aient une confiance accrue dans le système [DB01, FRP11]. Les cliniciens accepteront plus facilement les erreurs du système si elles sont accompagnées d'une incertitude et seront plus à même de répondre rapidement à une alarme de qualité produite par le système (à condition que ce type d'alarme ait une forte spécificité).

Dans le contexte de notre application, nous sommes confrontés à la gestion de l'incertitude et l'imprécision relatives aux données fournies par nos capteurs que sont les caméras puis la prise en compte de l'incomplétude des connaissances médicales dans la formalisation des scénarios médicaux décrits par un expert médical.

1.3. Le système de vidéo-surveillance multicaméras proposé en complément à l'existant en USI

La surveillance en USIs revêt un caractère très spécifique, important et diffère de toute autre forme de surveillance quel que soit leur domaine d'application par la complexité de ses données. L'incertitude, l'imprécision relative à l'information médicale et l'incomplétude des connaissances sont des aspects très importants à prendre en compte dans l'amélioration du système de monitoring en USI.

Les systèmes de surveillance en USIs actuels restent relativement très primitifs et alarmistes malgré les efforts considérables en termes de recherches depuis quelques années dans ce sens. Ces systèmes de surveillances en USIs comportent trop de fausses alarmes qui reflètent tout simplement le dépassement d'un seuil par le paramètre surveillé. Les alarmes ne correspondent pas toujours à la criticité de l'état du patient. Ce qui entrave la capacité des infirmières de réagir rapidement avec l'intervention clinique appropriée [NGK⁺12].

Les alarmes introduisent une pollution sonore constante et parfois gênent la quiétude des patients et constituent de ce fait un élément gênant pour la stabilité de l'état respiratoire du patient. Ce système de surveillance actuel en soins intensifs très alarmistes souffre d'un manque aigu d'information visuelle en temps réel des actions, des mouvements, des comportements et de l'état du patient ainsi que ceux des médecins pendant le séjour à l'hôpital.

En conséquence, la problématique d'amélioration de la performance des systèmes de surveillance en USIs est d'actualité depuis une décennie et représente une bonne part de thématique de recherche en matière d'équipements technologiques médicalisés dans les instituts de recherche. Elle concerne, pour la plupart, des travaux d'état de l'art en la réduction du taux de fausses alarmes et la mise en place de protocoles informatisés qui mettent en exergue des techniques automatiques et rapides de traitement de l'information en USIs pour aider les médecins surveillants dans l'exercice de leur fonction.

Dans nos travaux, nous abordons cette problématique d'amélioration du système de monitoring de façon différente. Des travaux significatifs avec de nombreuses améliorations ont été faits pour réduire le taux de fausses alarmes. Cependant, ils en

existent moins dans le cadre de l'enrichissement de nouvelles fonctionnalités qui rendront ces systèmes de surveillance en USIs plus intelligents et robustes afin que les salles de soins intensifs deviennent intelligentes et omniprésentes : smart-USI. Pour cela, nous proposons que les salles de soins soient dotées de deux grandes catégories de capteurs pour la surveillance des patients que sont des capteurs physiques permettant l'acquisition de paramètres physiologiques du patient (ils représentent l'existant en équipements de surveillance en USI actuel tels que l'ECG : électrocardiogramme) et des capteurs de vision (ils représentent un système de vision intelligent de perception multicaméras en complément des systèmes de surveillance actuel en USIs).

Les capteurs physiques fournissent des informations de types analogiques ou numériques sous forme discrète et qui représentent les seuils franchis par un paramètre physiologique surveillé, conséquence des alarmes sonores. Les capteurs de vision sont des caméras qui fournissent des informations de type visuel (Images/Vidéos) représentant des actions, des comportements, des états cliniques physiologiques des patients et aussi des scènes cliniques médicales (actions/activités du clinicien) en temps réel.

Cette approche est inspirée du fait que naturellement l'audition et la vision sont complémentaires et leur exploitation conjointe permet d'améliorer la qualité de l'information pour une meilleure prise de décision. De même, cette approche est aussi justifiée par le manque crucial de contenu visuel (images), de l'état clinique du patient et des actions du médecin dans les unités de soins intensifs. Globalement, le système de vidéo surveillance intelligente en unité de soins intensifs étant un outil d'aide à la décision pour les médecins surveillants et les cliniciens dans le processus de prise de décision [SDG05] possède deux principaux avantages que sont :

- la détection intelligente d'événements cliniques importants suivis d'alarmes visuelles : cette fonctionnalité permet d'aviser, sur la base de reconnaissance automatique d'un scénario intéressant par un enclenchement d'alarme visuel, les médecins surveillants par rapport à un état, comportement ou une action surveillée chez le patient.

- la sélection de vidéo intelligente basée sur la reconnaissance automatique pour la mise en place d'une boîte noire médicale qui contient un résumé des événements critiques ou importants pendant le séjour hospitalier d'un patient en unités de soins intensifs.

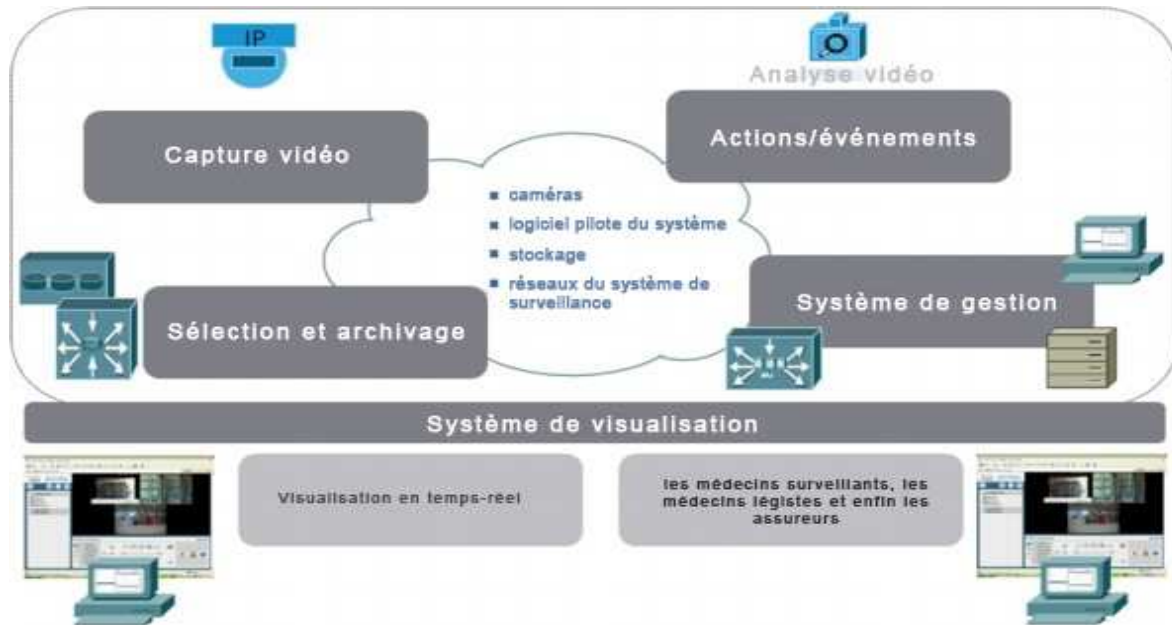


Figure 1.3. Architecture du système de traitement de vidéosurveillance multicaméras [AME⁺14].

1.3.1. Architecture physique et organisation multicaméras du système de vidéo surveillance en USIs

L'absence quasi-totale d'informations visuelles en temps-réel et d'alarmes moins pertinentes et significatives sous forme visuelle constatée par une étude et une enquête sur les systèmes de surveillance actuels en salles de soins intensifs nous a motivé à proposer un système de vidéo surveillance automatique et intelligent en complément au système alarmiste et sonore de monitoring des patients existant. Ce système de vidéo surveillance en USIs est doté d'une fonction principale qu'est la reconnaissance de scénario. Cette dernière étant définie comme une situation à risque pour le patient susceptible de modifier son état de santé dans les instants à suivre ou une action médicale (scènes cliniques) entreprise par le médecin.

La reconnaissance de scénario est basée sur un processus complexe en vision par ordinateur d'analyse et d'interprétation automatique de la scène pour la surveillance

des activités humaines en milieu médical. A cette fin, plusieurs modules spécialisés de vision sont nécessaires. Leurs rôles respectifs sont, la détection, le suivi des objets mobiles, la classification des objets et l'analyse des comportements afin de reconnaître des scénarios modélisés. L'application principale envisagée consiste à développer un système visuel de monitoring de Détection Automatique de Situations à risque et d'Alerte (DASA) pour la surveillance des patients en environnement médical. Une fois la phase de la reconnaissance et la détection du scénario effectuée, le système doit alerter le corps médical en amont puis effectuer l'archivage vidéo via un processus de sélection vidéo dans une base de données spécifique nommée Boîte Noire Médicale (BNM).

La seconde application envisagée dans ce cadre est l'étude, la conception avec la mise en œuvre d'un système de sélection vidéo intelligente et d'archivage automatique dans la Boîte Noire Médicale. Ces deux applications permettent d'améliorer de façon significative le monitoring des patients en salles de soins intensifs et, de ce fait, permet d'améliorer la qualité de services des soins médicaux en unités de soins intensifs. L'architecture physique du système est organisée autour de plusieurs caméras dont le nombre dépend de la configuration de la salle de soins (la superficie, la hauteur et du nombre de lits pour les patients). Les caméras sont calibrées pour couvrir la totalité de la scène dans la salle de soins et sont divisés en deux catégories celles (caméra à œil de Fish) fixées au plafond et celles qui font face aux lits des patients.

L'architecture physique du système est basée sur un réseau de caméras interconnectées via un support de type câble ou un réseau wifi si autorisé dans l'hôpital pour plus de flexibilité d'accès. Le système de vision à perception multi-caméras est une approche assez récente et qui reste en cours d'investigation dans plusieurs laboratoires de recherche en vision par ordinateur. Les nœuds du réseau de caméras échangent entre des informations pertinentes de haut niveau pour améliorer la fonction de reconnaissance effectuée par un serveur central. Mais les processus de vision de bas niveau sont effectués localement au niveau de chaque caméra dotée d'un firmware programmable dans lequel est embarqué une application basée sur des algorithmes de vision de traitement bas niveau tels que la détection, la classification et le suivi des objets mobiles (médecins, patients et visiteurs).

L'avantage principal de cette architecture est la distribution des fonctions locales d'interprétation bas niveau sur des caméras intelligentes dotées de capacités de traitements en temps-réel et de communication. Seules les informations de haut niveau transitent entre les nœuds de ce réseau. Les traitements bas niveau englobent la détection, le suivi local et la classification des objets.

Ce type d'architecture est particulièrement utile pour faire face aux contraintes temps-réel du système d'interprétation en ligne et plus particulièrement pour un système de vision à perception multi-caméras. Il permet aussi de gérer de façon plus élégante et avec une flexibilité les imperfections des données telles que l'incertitude et l'imprécision liées aux traitements bas niveau. Le système de vision à perception multicaméras proposé intègre plusieurs fonctionnalités selon une architecture centralisée et distribuée (Architecture Centro-Distribuée: ACD). Une description plus détaillée des fonctionnalités du système est proposée dans la section suivante qui est consacrée à la description de l'architecture logique du système.

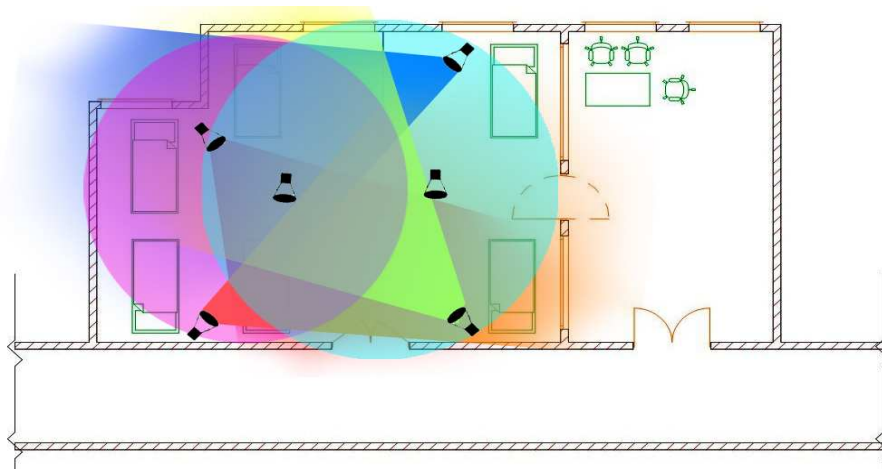


Figure 1.4. : Topologie physique du système de vidéosurveillance multicaméras [AME13].

Dans l'architecture du système expérimental installé aux unités de soins intensifs au service cardiologique, nous avons utilisé six caméras réparties comme suit : deux caméras PTZ à œil de poisson installées au plafond puis quatre autres installées face aux différents lits des patients. L'ensemble du système de vision possède des zones de recouvrement offrant de ce fait des possibilités de fusion d'informations pour améliorer la qualité de la décision du système.

1.3.2. Architecture logique du système de vidéo surveillance en USI

L'architecture logique du système de vidéosurveillance intelligente proposée en unités de soins intensifs exploite des modules de visions distribués. Le flux d'informations est séparé en deux catégories principales à savoir : le flux de Trafic de Base Vidéo (TBV) qui représente le flux multi média natif vidéo et le flux de Trafic de Commandes Vidéos (TCV) qui représente les informations à échanger entre capteurs d'acquisition vidéo (caméras). Seules les informations de haut niveau transitent entre les nœuds du réseau. L'ensemble des traitements est distribué via un système de perception de vision à architecture hybride Centro-Distribuée. D'une part, les traitements bas niveau englobent la détection, le suivi local et la classification des objets qui sont mis en œuvre à l'aide d'application embarquée dans une freeware intégrée dans une mémoire interne des caméras intelligentes et distribuées localement à travers les différentes caméras.

Au total, les caméras appartenant au réseau de surveillance échangeront des informations spécifiques de commandes entre eux en exploitant les canaux d'échanges de type TCV afin de coopérer et collaborer pour améliorer l'analyse et l'interprétation de la scène vidéo.

1.4. Vers une Boîte Noire Médicale en USIs

Au regard de l'avancée considérable de la technologie d'instrumentation et plus particulièrement dans le domaine médical et spécifiquement dans les unités de soins intensifs, le besoin de la création d'une boîte noire médicale à l'image de la boîte noire d'un avion se fait de plus en plus ressentir. Elle constitue un véritable défi et une idée originale très récente que nous proposons et est d'une grande utilité pour les médecins. Tous les praticiens hospitaliers accordent tous leurs violons sur la grande utilité qu'elle constitue et qui n'est plus à démontrer. La conception et la mise en place d'un tel outil sont motivées par la nécessité de créer un outil d'aide au diagnostic et à la décision médicale.

Du système de vidéo surveillance intelligent proposé en couplage au système de monitoring existant en USI, d'énormes collections de vidéos médicales sont disponibles avec plusieurs applications pour divers services médicaux tels que l'imagerie médicale, des diagnostics médicaux, la formation des professionnels de la santé, l'aide à la décision pour les médecins légistes, une source de preuve pour les assureurs de santé, la recherche médicale et de l'éducation en santé.

Le système de surveillance vidéo intelligente permet de gérer, d'analyser, d'indexer et de récupérer les informations de flux vidéo IP. Après l'indexation et l'extraction des séquences d'images clés en tant que bon représentant de scénarios importants. Ce système permet de faire le résumé des événements d'intérêts lors d'un séjour hospitalier. La boîte noire médicale est une base de données contenant des images intéressantes obtenues par sélection automatique et intelligente de vidéo sous forme de résumés d'événements critiques ou importants d'un séjour hospitalier d'un patient.

Le concept de la boîte noire médicale est une nouvelle idée originale qui présente un double intérêt : qu'il s'agisse d'une assistance ou d'une aide à la décision pour la surveillance médicale dans les unités de soins intensifs. Elle est nécessaire et constitue une aide au diagnostic médical après interrogation de la base multimédias suivi de l'analyse du contenu vidéo de la boîte noire médicale. En conséquence, elle peut contribuer au développement de nouveaux protocoles d'intervention qui permettront d'améliorer les performances des soins médicaux en salles de soins intensifs. L'architecture de la base de données, les techniques d'indexation et la recherche d'informations du contenu multimédias que constitue la boîte noire médicale est un nouvel axe de recherche que nous aurons à aborder dans nos futurs travaux.

L'exploration de ces bases de vidéo aide aussi à l'élaboration de nouvelles connaissances médicales et la construction de nouveau scénario à reconnaître. La boîte noire médicale reste dans tous les cas une solution très prometteuse pour l'amélioration de la performance des soins médicaux afin d'améliorer la qualité de service médical (Medical-QoS) en milieu hospitalier ceci pour une meilleure prise en charge des patients.

Conclusion

Nous avons présenté dans ce chapitre, les systèmes de vidéosurveillance intelligents, leurs enjeux, leurs fonctionnalités et leurs principales applications ont été présentées. L'organisation et les composantes d'une salle de soins intensifs ont été présentées. De même, la surveillance en unités de soins intensifs a été présentée et nous avons montré sa spécificité par rapport aux autres types de surveillance. Cependant, les limites de la surveillance actuelle ont été mises en évidence et nous pouvons noter :

- surveillance alarmiste et présence d'un taux de fausses alarmes considérables qui gênent la quiétude du patient. De plus, des alarmes moins pertinentes et moins informatives ne favorisent pas le monitoring;
- une surveillance très complexe du fait de la nature imprécise et incertaine des données qui s'y trouvent avec une incomplétude des connaissances du domaine médical. En conséquence des événements médicaux difficilement formalisables ;
- un manque de données visuelles pertinentes décrivant les comportements et les états du patient.

Pour répondre à ces limites de la surveillance en USI, nous avons introduit un système de vidéo surveillance automatique et intelligent pour suppléer le système existant qui a été décrit tant bien via son architecture physique que logique. Enfin, un nouveau concept et novateur de la création d'une Boîte Noire Médicale (BNM) a été présenté comme une solution très prometteuse afin d'aider à une phase exploratoire pour la représentation des connaissances incomplètes (par exemple les scénarios non décrits par le médecin).

Chapitre 2

Etat de l'art des modules de traitements d'un système de vision pour la reconnaissance de scénario

Introduction

Les systèmes de reconnaissance automatique d'activités exploitent l'analyse et l'interprétation vidéo. L'analyse et l'interprétation vidéo consistent à reproduire, à l'aide d'un ordinateur, l'analyse qu'effectuerait un humain en regardant les séquences vidéo provenant des caméras de surveillance. Ce chapitre expose un état de l'art des différents traitements de base pour la reconnaissance de scénarios dans des séquences vidéo en réalisant un processus d'analyse et d'interprétation vidéo. En général, un tel système est basé d'une part, sur des traitements bas niveau (correction, amélioration et formatage de l'image) et d'autre part, sur des traitements haut niveau (techniques d'analyse et d'interprétation) pour transformer l'observation ou la mesure que représente l'image (signal) en des décisions (enchaînement de symboles) tout en y intégrant une gestion d'incertitude et d'imprécision des processus de vision et des données traitées. La Figure 2.1 illustre les différentes grandes étapes d'une architecture organisée en niveau (bas et haut) d'un système de vision pour la reconnaissance de scénarios.

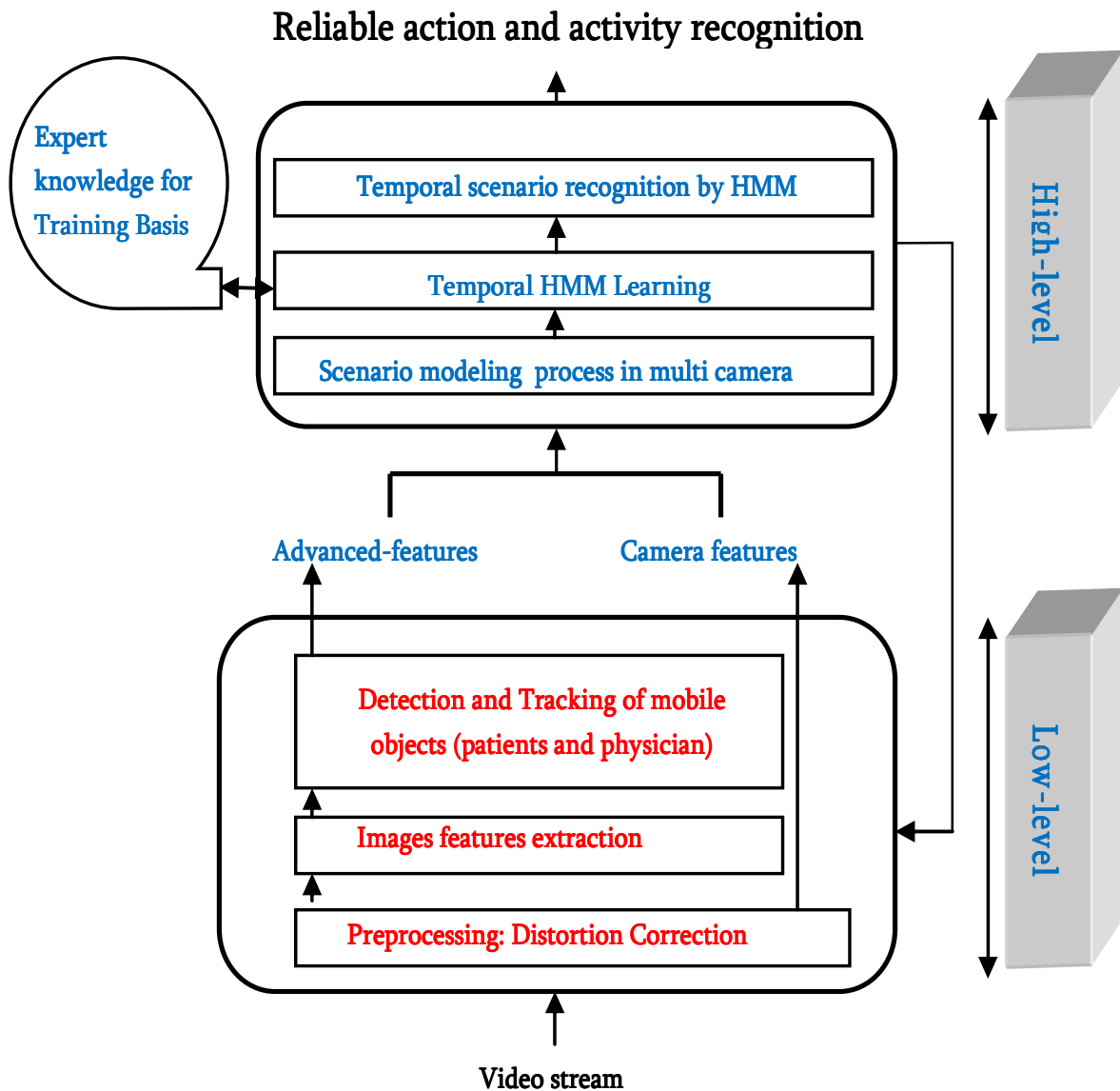


Figure 2.1. Architecture en couche modulaire d'un système de vision pour la reconnaissance de scénario dans une séquence vidéo [AME14].

De manière générale, la reconnaissance de scénarios dans des séquences d'images pour la vidéosurveillance automatique concerne la caractérisation des mouvements d'objets présents dans la scène. C'est un processus de vision par ordinateur très complexe qui impose la création de plusieurs modules spécialisés que nous pouvons classer en traitements bas niveaux (prétraitement, détection et classification, et suivi d'objets de la scène) et en traitements haut niveau qui concerne l'interprétation haut niveau de la scène en se basant sur les sorties de l'étage précédent

et des outils mathématiques basés sur des techniques symboliques ou orientés numériques.

2.1. L'acquisition et le prétraitement d'images

En vision par ordinateur, les deux grandes étapes que compose le processus de vision sont : l'acquisition des images et leurs traitements proprement dits. A ces deux principales étapes s'ajoutent une étape intermédiaire «*le prétraitement*» très important et nécessaire en fonction du domaine d'application. L'efficacité des algorithmes des traitements proprement dits dépend aussi de cette étape intermédiaire.

Dans cette section, nous nous intéresserons de façon succincte aux notions relatives à l'acquisition et le prétraitement des données. Nous reviendrons de façon plus détaillée en présentant l'approche de prétraitement proposée pour la correction de distorsion d'images médicales tout en faisant une analyse des différents résultats obtenus dans le chapitre 5.

2.1.1. L'acquisition de séquences d'images vidéo

L'acquisition permet de convertir avec un capteur (caméra), une vue réelle en une image qui peut être numérisée selon une chaîne. Le système d'acquisition est un ensemble de phénomènes intervenant au cours du processus de formation des images. Le processus d'acquisition se fait via une technique de formation géométrique des images. La façon la plus simple d'expliquer le processus de formation d'une image (l'acquisition), passe par la présentation de l'appareil sténopé.

L'appellation anglophone de cet appareil est «*pinhole camera*» et permet aisément d'imaginer le processus de formation d'une image. En effet, cet appareil est une boîte hermétique à la lumière, dans laquelle est percé un trou dont le diamètre est supposé nul. Ainsi, tous les rayons lumineux passant par le trou projettent la scène observée renversée sur le fond de la boîte comme illustré à la Figure 2.2.

Nous définissons le modèle géométrique d'une caméra en considérant deux transformations: une projection perspective qui transforme le point 3D de l'espace en un point 2D dans le repère caméra et une transformation caméra image qui permet

d'établir le passage entre le repère caméra et le repère image. La Figure 2.2 en est une illustration.

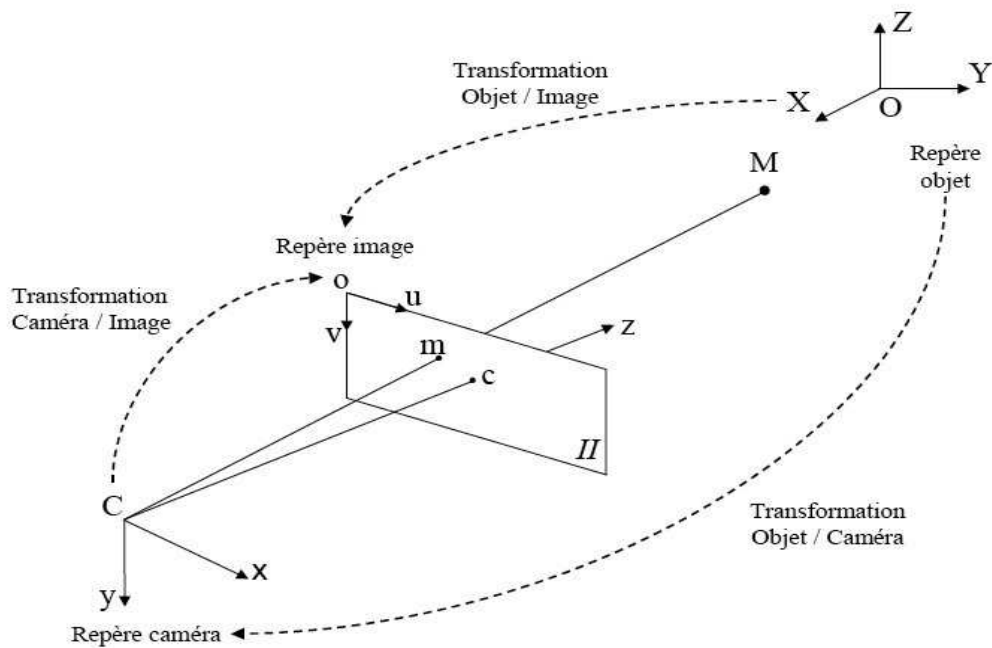


Figure 2.2. Modèle géométrique sténopé d'une caméra [AME12].

2.1.2. Calibrage et paramètres d'une caméra

La calibration a pour objectif d'estimer les paramètres de la caméra afin de disposer d'une connaissance sur la géométrie du système de vision. On distingue deux principaux types de paramètres : les paramètres intrinsèques à la caméra et les paramètres extrinsèques. Les paramètres intrinsèques sont ceux qui traduisent le passage du repère caméra au repère image [Zha00]. Ils sont sous la forme d'une matrice appelée matrice caméra (cette matrice nommée C est estimée en utilisant l'équation 2.1). Elle est composée d'éléments qui représentent des paramètres fixes, inhérents à la conception de la caméra et de son optique (ils peuvent être modifiables dans le cas d'une caméra à focal réglable).

$$C = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

Le couple (u_0, v_0) représente les coordonnées (x, y) du point principal dans le repère image. Par ailleurs, en notant f la distance focale, on a $\alpha_u = -k_u \cdot f$ et $\alpha_v = k_v \cdot f$ où k_u et k_v sont les facteurs d'échelle respectivement vertical et horizontal. Notons que les coordonnées sont exprimées en pixels et les coefficients d'échelle en pixels/mm².

Les paramètres extrinsèques sont les paramètres qui lient le repère caméra avec le repère de l'objet de calibration (donc le repère monde). Ces paramètres varient en fonction de la position du capteur dans l'espace. Ils sont au nombre de douze et forment une matrice (équation 2.2) de rotation et un vecteur de translation permettant d'effectuer la transformation du repère caméra vers le repère monde [KBS00] :

$$A = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

L'obtention d'une image analogique $e(x, y)$ dans le plan image d'un système optique est toujours limitée, car tout système optique transforme un point objet en une image sous forme de tache. Plusieurs raisons peuvent expliquer l'apparition de cette tache [GD00] :

- le défaut de mise au point si l'image n'est pas exactement formée dans le plan image ;
- les aberrations chromatiques si la lumière n'est pas assez monochromatique ;
- les aberrations géométriques si les dimensions transversales du faisceau lumineux sont trop grandes, ou diffraction de la lumière par les diaphragmes et ouvertures du système optique.

Contrairement aux aberrations géométriques, la diffraction est d'autant plus importante que les dimensions transversales du faisceau lumineux sont petites, en d'autres termes, si les diaphragmes du système sont petits. La diffraction est en général la contrainte la plus importante. En outre, les optiques réelles ne se comportent pas exactement comme le modèle affine ou sténopé, mais présentent des aberrations. Le modèle de caméra décrit jusqu'à présent peut être amélioré en tenant compte des distorsions linéaires ou non. Des distorsions de trois natures différentes peuvent être considérées [Bou00] :

- la distorsion radiale due le plus souvent à des défauts dans la courbure radiale des lentilles de la caméra ;
- la distorsion de décentrage due à un mauvais alignement des centres optiques des lentilles ;
- la distorsion prismatique due à une inclinaison des lentilles les unes par rapport aux autres.

Dans le cadre de l'application de vidéosurveillance intelligente que nous développons en USI, nous avons utilisé des caméras à œil de poisson du constructeur AXIS installées au plafond de la salle de soins qui fournissent des séquences d'images en y introduisant une distorsion radiale. Compte tenu de la sensibilité du domaine d'application qu'est le domaine médical (la santé ou la vie humaine), il convient de faire un prétraitement sur ces images avant de les analyser pour l'étage supérieur d'interprétation haut niveau de la scène. L'algorithme proposé est présenté avec les résultats obtenus et leurs analyses au chapitre 5. Dans la section suivante, nous présentons le prétraitement des images.

2.1.3. Le prétraitement des images vidéo acquises

Le «*prétraitement*» (ou avant traitement) est une importante étape intermédiaire entre la phase d'acquisition des séquences d'images. Il offre des traitements permettant de corriger les différentes aberrations introduites sur les images pendant la phase d'acquisition afin de disposer de données images proches du réel. Le prétraitement désigne l'ensemble des opérations qui ont pour rôle de donner à l'utilisateur une image libérée des défauts propres au(x) capteur(s). Cette étape peut en fait être très difficilement déconnectée des étapes ultérieures du traitement.

Le prétraitement permet de corriger, d'améliorer et de restaurer l'image. Il permet aussi de formater l'image. De plus, il intègre un certain nombre de processus permettant de préparer l'image au traitement de couche supérieure notamment l'analyse (décision ou interprétation) de l'image. L'amélioration a pour but de satisfaire généralement l'œil de l'observateur humain. L'œil humain étant essentiellement sensible aux forts contrastes, il faut par conséquent trouver des moyens pour améliorer l'image en essayant d'augmenter le contraste dans le but d'accroître la séparabilité des régions

composant une scène. Pour ce faire, deux principales variantes de techniques sont utilisées : les techniques basées sur l'histogramme de couleur et les techniques basées sur le filtrage fréquentiel.

D'autres approches existent et peuvent être aussi utilisées. La compression est une opération de formatage, ou en d'autres termes c'est une adaptation de la donnée dans un format plus portable. Deux familles d'algorithmes de compression existent : la compression sans perte et la compression avec perte. La restauration d'images est un processus de correction d'images bruitées (cas d'une distorsion). Elle a pour objet la réduction et l'élimination des distorsions introduites (bruits) par le système ayant servi à l'acquisition de l'image. Son but est d'obtenir une image qui soit la plus proche possible de l'image idéale qui aurait été obtenue si le système d'acquisition était parfait. La technique la plus utilisée reste le filtrage. Ce dernier peut être en fonction du domaine de traitement du signal (image) temporel ou fréquentiel puis en fonction du modèle d'état, il peut être linéaire ou non linéaire.

Les sources de bruit lors de la phase d'acquisition sont multiples. Nous avons le bruit lié au contexte de l'acquisition (mouvement du capteur), à la dérive lumineuse puis le bruit lié au capteur qui peut être une distorsion de la gamme des niveaux de gris, distorsion géométrique, mauvaise mise au point et enfin un bruit lié à la numérisation (échantillonnage, quantification, codage). Dans notre application, nous avons proposé et exploité une méthode de correction de distorsion de type radiale via une approche de correction de distorsion d'images par auto calibration stéréoscopique à partir de scène multi vue. Nous regroupons en deux principales classes les différents types de correction que sont les corrections photométriques et les corrections géométriques.

2.1.3.1. Les corrections photométriques

Elles concernent les défauts de linéarité et les défauts d'homogénéité. Les défauts de linéarité sont assez facilement corrigés par des tables de transcodage qui sont rapides. Les valeurs du signal corrigé n'étant pas calculées mais directement recherchées en adressant une mémoire précalculée appelé LUT. Les défauts de linéarité nécessitent un calibrage des organes et une analyse qui se fait soit sur une rampe de gris

continue, soit sur quelques points fixes choisis. Le meilleur exemple de ce type de correction est le re-calibrage (ou correction) Gamma d'un écran cathodique.

En effet, un écran CRT (*Cathode Ray Tube*) n'est pas un système électronique linéaire, ou plus simplement, l'intensité lumineuse reproduite à l'écran n'est pas une fonction linéaire de la tension d'entrée. D'un point de vue strictement physique, on pourrait s'attendre à ce qu'une correction gamma soit le processus qui suffise à compenser la non linéarité entre la tension d'entrée et la lumière affichée à l'écran pour obtenir une correcte reproduction de l'intensité lumineuse en un point de l'écran. Cependant, la perception humaine de la luminosité n'est pas elle même uniforme mais se comporte plutôt comme une fonction puissance de l'intensité. Finalement pour obtenir une bonne correction gamma, il faut connaître les caractéristiques physiques de l'écran, le comportement de la vision humaine (ceci oblige à définir une correction gamma pour un observateur de référence avec des conditions d'observation standard) et les conditions de prise de vue des images.

Les défauts d'homogénéité sont dus aux inhomogénéités spatiales des capteurs ou éventuellement aux défauts des optiques ou des éclairages. Leur correction rigoureuse est très coûteuse et donc rarement effectuée. De ce fait, il est préférable d'effectuer des corrections approximatives, soit par segmentation de l'image en quelques zones quasi-homogènes, soit en appliquant des corrections locales (soustraction d'une moyenne flottante, division par un contraste flottant, etc.), soit en appliquant des corrections globales approchées (approximation par des quadriques par exemple). L'assurance d'avoir des documents numériques présentant une très bonne homogénéité de la photométrie permet de gagner en temps de calcul dans les étapes ultérieures.

2.1.3.2. Les corrections géométriques

A contrario des corrections de linéarité, les corrections géométriques sont évidemment spatialement variantes, et de ce fait sont difficiles pour la généralisation. Si nous appelons (x,y) les coordonnées d'un point dans une image de référence (non échantillonnée) et (u,v) les coordonnées de ce même point après déformation

géométrique dans l'image distordue, la correction géométrique se caractérise par la transformation G qui associe le couple (u,v) au couple (x,y) tel que :

$$(x, y) = G(u, v) \quad (2.3)$$

On appelle correction géométrique inverse la transformation G_{inv} qui associe (x,y) à (u,v) telle que :

$$(u, v) = G_{inv}(x, y) \quad (2.4)$$

Dans le cas d'une image continue, la valeur $f(x,y)$ de l'image de référence au point (x,y) est égale après distorsion à la valeur $f'(u,v)$ de l'image distordue au point (u,v) . En traitement numérique, les images sont généralement échantillonnées sur un maillage rectangulaire. Nous ne connaissons le signal $f'(u,v)$ de l'image distordue que sur un réseau discret $L_{u,v}$ et nous nous intéressons au signal $f(x,y)$ de l'image de référence que sur un réseau discret $L_{x,y}$. Les pixels appartenant au réseau $L_{x,y}$ ne correspondent pas aux pixels appartenant au réseau $L_{u,v}$. Il convient donc de déterminer les bonnes valeurs par interpolation des pixels environnants. Ceci peut être fait selon deux méthodes différentes :

- Dans la première, partant d'un pixel (x_0, y_0) de $L_{x,y}$, on calcule (u_0, v_0) , puis on détermine la valeur de $f(x_0, y_0)$ en interpolant la valeur équivalente de $f'(u_0, v_0)$ à l'aide des valeurs connues de l'image distordue $f'(u, v)$ sur le réseau $L_{u,v}$.
- Dans la deuxième, on transforme chaque pixel (u, v) de $L_{u,v}$ qui permet d'obtenir des pixels corrigés (x, y) extérieurs à $L_{x,y}$, de valeur $f(x, y) = f'(u, v)$ connues. Puis, on calcule la valeur $f(x_1, y_1)$ d'un point du réseau $L_{x,y}$ en interpolant les valeurs $f(x, y)$ des points corrigés environnants.

Les corrections géométriques ont une grande importance pour les images médicales du fait de la sensibilité du domaine où les aberrations sont notables si les capteurs présentent des défauts inhérents à leur constitution. L'utilisation des caméras à œil de poisson pour l'acquisition des images médicales constitue une cause de distorsion géométrique. Nous exposerons brièvement dans les annexes les trois principales types

de corrections présentant une complexité croissante en fonction des techniques de corrections qui sont utilisées.

2.2. Les traitements bas niveau

La chaîne de traitement vidéo (séquences d'images) en vision peut être décomposée en deux niveaux: «*bas-niveau*» qui se construit autour des méthodes d'analyse d'image ayant pour but d'extraire des caractéristiques telles que les pixels, les régions d'intérêts, contours, texture des images analysées sans les interpréter et le «*haut-niveau*» qui intègre l'ensemble des méthodes permettant d'interpréter les caractéristiques issues du bas-niveau dans le cadre d'une prise de décision, de la classification, de la reconnaissance de forme et de l'analyse et de l'interprétation vidéo par exemple. En d'autres termes, le bas niveau qui consiste à appliquer des traitements à base de filtrage et un haut niveau qui permet de prendre une décision fiable.

Le bas niveau permet de définir des jeux de caractéristiques destinées à alimenter le niveau supérieur (le haut niveau) à des fins d'analyse, d'interprétation et de reconnaissance par exemple. Les tâches de bas niveau sont souvent basées sur des filtrages, en particulier des corrélations. Ils permettent généralement de faciliter les traitements «*haut-niveau*» qui sont relativement plus complexes et qui sont nécessairement exploités pour aboutir à des décisions fiables (théorie de fusion de l'information ou intelligence artificielle).

Les études sur la partie «*bas-niveau*» se trouvent appliquées dans des opérations aussi variées que la compression et le cryptage de l'information (images) ainsi que dans la détection et le suivi d'objets mobiles pour une chaîne d'analyse et d'interprétation de séquences vidéo. Ces traitements ont pour but d'extraire l'information utile et pertinente contenue dans l'image au regard de l'application considérée. Ces traitements opèrent sur des données de nature numérique et doivent donc simplifier l'image sans trop la dégrader. Le résultat d'un tel traitement donne une description structurelle de l'image sans qu'il y ait de lien avec le contexte de la scène réelle. L'un des principaux outils utilisés est par exemple la segmentation d'images.

Dans cette section, nous nous concentrons plutôt sur les traitements «*bas-niveau*» de séquences d'images (vidéo) dans un contexte d'interprétation automatique de scénario dans la scène vidéo.

2.2.1. La détection d'objets mobiles dans une vidéo

Une littérature bien fournie, assez complète, existe et fait état de nombreuses approches en termes d'algorithmes de détection et de suivi d'objets mobiles dans une vidéo [Mot07]. Le module de détection du mouvement permet d'extraire à partir d'une séquence d'images, les zones contenant des objets se déplaçant dans la scène observée [Nag83]. Pour cela, il faut réaliser une projection de la région occupée par cet objet dans le plan image afin de former un amât de pixels d'une image spécifique appelée le masque. Ce masque peut constituer une zone d'intérêt de l'image lors de la phase d'analyse et d'interprétation de la scène. En utilisant un critère de classification basé sur le taux d'utilisation de ces algorithmes, nous pouvons les grouper en deux catégories.

La première catégorie concerne les algorithmes très répandus qui sont des méthodes classiques telles que les approches basées sur l'appariement entre images consécutives. La seconde catégorie est celles des algorithmes moins répandus mais très robustes et relativement assez récents telles que les approches basées sur une modélisation du fond. S'agissant des algorithmes les plus répandus, l'algorithme le plus connu, simple et direct réalisant la détection de mouvement entre deux images utilise une différence d'images pixel à pixel [Nagel83]. Les pixels de l'image différence dont le niveau de gris dépasse un certain seuil sont considérés affectés par le mouvement.

Cependant, dans le cas de scènes réelles, diverses perturbations rendent la décision de détection difficile. En effet, dans les scènes d'intérieur ou d'extérieur, en dehors des déplacements réels des objets, il existe des phénomènes dynamiques d'origines multiples qui affectent la scène elle-même ou les images de la scène [MVB07]. Les origines de ces changements peuvent être classées selon trois types de phénomènes que sont :

- les changements de luminosité et les ombres ;
- la vibration de la caméra ;
- le bruit d'acquisition et du capteur.

L'objectif d'un module de détection idéal pour les tâches de surveillance est alors de mettre en valeur uniquement les masques des objets en mouvement utiles à l'analyse en éliminant l'ensemble des perturbations affectant l'image. La démarche est soit une modélisation de manière explicite de l'image afin de les filtrer ultérieurement, ou soit une normalisation des données brutes pour les rendre insensibles aux perturbations. La décision de détection du mouvement est réalisée par le biais d'une analyse de la variation temporelle des observations. Cette différence peut être effectuée entre images consécutives (inter-images) ou entre l'image courante et une image de référence exempte d'objets en mouvement.

La première approche est plus simple dans le sens où elle ne requiert pas d'image de référence. La difficulté de cette approche est la détection du masque complet de l'objet lorsque les masques d'un objet déterminés sur deux images consécutives présentent une intersection importante. En particulier cette stratégie ne permet pas de détecter un objet dès qu'il devient immobile.

2.2.1.1. Technique basée sur la différence entre deux images consécutives

Cette technique effectue une soustraction entre deux images successives. En effet, il s'agit de soustraire une image acquise à un instant t d'une autre acquise à un instant $t+1$. Ce faisant, les pixels de l'image résultante dont l'intensité est proche de zéro sont considérés comme étant les pixels de fond. Cette technique a pour avantage de consommer moins de ressource informatique pour sa mise en œuvre et sa flexibilité car elle ne nécessite pas une phase d'initialisation à base de scène statique. Par contre, le choix de seuil de décision (proche de zéro) qui peut être local ou global constitue son point de vulnérabilité. Ce seuil pour être robuste doit aussi tenir compte des phénomènes pouvant affecter la qualité de la scène comme la luminosité, le bruit et le mouvement du capteur. Le résultat de la soustraction entre deux images consécutives est la décision si des objets présents dans la scène ont bougé. Pour cela, il faut choisir un attribut sur lequel la décision est prise.

La décision est généralement effectuée sur un pixel ou une région. Dans le cas où la décision s'effectue sur un pixel avec un seuillage global, on décide si un pixel est en mouvement ou pas en comparant ces coordonnées dans les deux images consécutives et

l'ensemble des pixels détectés en mouvement forme la carte binaire de mouvement. Il existe d'autres techniques qui permettent de décider, d'une part, à partir d'un rapport de vraisemblance sur la différence d'intensité ou en utilisant des modèles statistiques plus complexes [Ros98] et, d'autre part, sur des tests d'hypothèses basés sur une modélisation de la différence d'image successive sous forme de mélange gaussienne avec l'hypothèse d'absence de mouvement et par une loi uniforme sous l'hypothèse de mouvement [Kon00].

Une décision basée sur région est effectuée par exploitation des modèles à contrario afin de détecter les régions de variation dans une séquence d'images. Le modèle à contrario détecte les fortes corrélations des différences d'orientation des gradients au niveau des zones de changements. La détection est effectuée sur des régions carrées de taille dynamique et variable déplacée sur toute l'image. Dans le cadre de la décision basée sur région, une approche a été proposée dans [LM03] pour détecter des changements dans des images satellitaires préalablement recalées. Dans [TVB04], la décision est prise sur des régions fournies soit par un découpage en blocs, soit par une segmentation basée sur les niveaux de gris. Cette approche a permis de faire une analyse des mesures locales de changement à partir des niveaux d'intensités de trois images successives.

2.2.1.2. Technique basée sur la soustraction de fond

Les techniques basées sur la soustraction de fond représentent les plus couramment utilisées. Elles nécessitent un apprentissage a priori du modèle du fond de la scène. De même, une phase d'initialisation est nécessaire et se fait en utilisant une image de référence générée en considérant que la scène observée ne contient aucun des objets qui seront détectés comme mobiles dans la suite. L'image de référence peut être générée en moyennant les N premières images de la séquence. Elle doit être mise à jour afin de s'adapter aux changements d'illumination et autres qui peuvent apparaître dans la scène observée. La conséquence de cette mise à jour est qu'elle peut entraîner la disparition d'objets restant immobiles pendant un temps assez court dans la scène.

La solution à ce problème se trouve dans l'apprentissage automatique des paramètres du modèle de fond dès le départ. Cette capacité d'apprentissage et

d'adaptation de ces techniques constitue, un des points forts qui rend l'approche riche et robuste face aux problèmes d'illumination. Mais, il convient de noter que ces techniques peuvent être regroupées en deux grandes catégories: la soustraction du fond par modélisation statistique et la modélisation du fond par moyenne récursive [Mot07].

La soustraction de fond par modélisation statistique est divisée en deux principales étapes : l'apprentissage du fond et l'étiquetage binaire des pixels. La distribution statistique en chaque pixel peut être représentée par une simple gaussienne pour un fond statique et un mélange gaussien pour un fond dynamique. En revanche, la soustraction par moyenne récursive se calcule en utilisant le paramètre alpha (α) inférieur à 1 qui est le paramètre d'oubli. Ce paramètre tend vers 1, lorsque la moyenne récursive tend vers l'image courante et vers 0, lorsque la moyenne récursive tend vers la moyenne arithmétique. Elle constitue l'approche que nous nous proposons d'exploiter dans le cadre de notre système d'interprétation des scènes médicales pour la vidéosurveillance intelligente en milieu hospitalier.

2.2.1.3. Technique basée sur le flot optique

La technique du flot optique permet de représenter par un champ de vecteurs les déplacements perceptibles dans une séquence d'images et fait l'hypothèse d'une caméra fixe. L'estimation du flot optique est faite sur l'hypothèse de conservation d'énergie et de l'énergie lumineuse qui est une hypothèse forte (obligatoire) pour une bonne estimation des vecteurs. Le flot optique constitue une information plus riche qu'une détection du mouvement ; elle peut servir dans un système de vision pour l'analyse des mouvements (par exemple : arrêt, marche, course, etc.) d'objets haut niveau de la vidéo. En détection de mouvement, les pixels appartenant au fond auront une vitesse nulle contrairement aux pixels qui font partie des objets mobiles. En général, la segmentation de mouvement repose sur l'association de chaque pixel au modèle de mouvement qui correspond le plus. Il s'agit de trouver des masques pour chaque objet en mouvement à condition d'avoir une différence de mouvement entre ces objets.

Dans [DG95], un champ de mouvement dense (par exemple le flot optique) est estimé et segmenté pour mettre en évidence des mouvements indépendants. De même, dans [Wix00], des mesures de flot optique cohérentes en direction sont accumulées

pendant quelques pas de temps. Par conséquent, l'estimation du déplacement de chaque pixel au long d'une séquence est obtenue. Cette méthode a été légèrement améliorée dans [TH05] pour fonctionner en temps réel. Le coût de calcul en termes de complexité algorithmique constitue une limite majeure de la technique du flot optique.

2.2.2. Le suivi d'objets détectés mobiles dans une vidéo

Nous présentons dans cette sous-section un état de l'art sur les principales techniques de détection d'objets en mouvement dans une vidéo d'une part et d'autre part sur les principales techniques de suivi «*tracking*» d'objets mobiles. En effet, suivre un objet mobile dans une scène à partir d'une caméra unique, consiste à utiliser les informations contenues dans une séquence d'images pour identifier la nouvelle position 2D (dans le plan image) ou 3D (dans l'espace) de l'objet suivi.

En vision par ordinateur, le suivi peut être considéré comme un problème d'appariement. Le suivi de personnes est fondamental dans les systèmes de vidéosurveillance puisqu'il est la base de l'analyse de comportements, la reconnaissance des activités et la détection des événements d'intérêts. Toutefois, le suivi est conditionné par la qualité de l'étape précédente qu'est celle de la détection car une mauvaise détection d'objets mobiles ne pourrait enclencher le suivi de cette dernière. Les techniques de suivi sont issues des « radaristes » pour le suivi de cibles, à base de théorie de l'estimation.

L'objectif de ces techniques étant de combiner les informations issues de divers capteurs et d'obtenir un état estimé le plus proche possible de l'état observé [JMC08]. Le filtre de Kalman est un exemple de techniques de trajectographie qui a connu beaucoup de succès et d'applications avec ces extensions. Le but des méthodes de suivi est d'estimer au fil du temps des paramètres d'une cible (ou plusieurs) présente dans le champ de vision de la caméra et initialement détectée par un moyen quelconque. Les paramètres peuvent être divers : position dans l'image, à laquelle peuvent s'ajouter la taille et l'orientation apparente, l'attitude, l'apparence, etc. Pour qu'une méthode de suivi soit complètement automatique, il faut non seulement suivre la ou les cibles mais aussi les initialiser automatiquement par une méthode de détection et gérer leurs éventuels arrêts et/ou sorties du champ de la caméra [Bug07].

Nous pouvons grouper en deux classes les techniques de suivi : celles déterministes et celles probabilistes. Les techniques probabilistes sont soit bayésiennes soit non Bayésiennes. Il existe d'autres approches de suivi qui exploitent une technique de segmentation dynamique qui sont utilisées lorsque l'on souhaite extraire la silhouette de la cible à chaque instant, et ce sans connaissance a priori sur sa forme. Basées sur une technique de segmentation successives d'images, les approches font généralement évoluer le contour de l'objet à l'instant précédent jusqu'à sa nouvelle position à l'instant courant. Un modèle d'état ou une minimisation des fonctions d'énergie sur le contour permet de faire un tel suivi.

Un bon état de l'art de ces techniques est fait dans le projet Canada [JMC08], ensuite la thèse de Aurelie Bugeau [Bug07] et enfin l'article de Yilamaz et al. [YJS06] qui traite de l'état de l'art des techniques de «*tracking*». Des topologies récentes de système de vision mettent en place une approche de suivi multicapteurs. Un système de suivi d'objets mobiles opérant à base de plusieurs capteurs a l'avantage d'élargir son champ d'action soit en exploitant la redondance des capteurs par combinaison ou bien par fusion d'informations pour améliorer la décision de suivi. Ces topologies multicapteurs engendrent un niveau de mise en correspondance supplémentaire. En effet, la fonction de suivi doit également être reliée par les observations des différents capteurs. Dans le cadre de notre application, nous avons utilisé une configuration multi capteurs pour le suivi des objets mobiles.

2.2.2.1. Techniques déterministes de suivi d'objets mobiles

La finalité d'un module de suivi d'un objet mobile est d'obtenir la trajectoire de cet objet en déplacement tout le long d'une séquence tout en gérant les problèmes d'occultations (on peut faire recours à un graphe treillis), ressemblance et de ré-identification et de fausse détection. Un objet à l'instant précédent est relié à un petit nombre d'observations à l'instant courant sur la base d'une mesure de distance. Les liaisons entre observations et objets sont pondérées par la distance associée à cette mesure de similarité.

Généralement, les techniques exploitent les caractéristiques des objets mobiles à base d'hypothèse de proximité (déplacement) et de modèle d'apparence telles que la

similarité de forme, le mouvement et le contenu photométrique. L'ensemble des caractéristiques peuvent être exploités de façon distincte, combinée deux à deux ou toutes ensembles en fonction de la complexité du problème de suivi à résoudre. A partir des caractéristiques de l'objet, les techniques déterministes permettent de mettre en correspondance des objets avec des observations entre deux images consécutives. Le but est, d'affecter les observations aux pistes en cours de suivi en minimisant la distance calculée entre les différentes caractéristiques. Le calcul de la distance entre les objets suivis et les observations sont basés sur des techniques de corrélation sur les contours de l'objet qui peuvent être normalisés au besoin pour faire face aux problèmes d'illumination. Il se peut que les caractéristiques de l'objet exploitées pour le suivi subissent des variations. Dans ce cas, les histogrammes de couleur ou de contour sont utilisés pour limiter les conséquences de ces derniers. La distance de Bhattacharya ou la divergence de *Kullback-Leibler* ou le flot optique [SA04] peut être utilisé lors de la mise en correspondance des objets avec deux observations entre deux images consécutives. Dans certains cas, le suivi d'un objet peut être étendu au suivi multi-objets en rendant interdépendantes les extractions des pistes.

Dans [FPK05], une technique est proposée et consiste à extraire dans un premier temps le meilleur chemin, ensuite modifier le graphe afin d'augmenter le coût des pistes passant sur les nœuds de ce chemin. Une fois le second chemin est obtenu, la procédure est reprise et ainsi de suite. Une gestion en ligne des occultations n'est pas possible mais est envisageable hors ligne lorsque l'objet est préalablement suivi avec cette approche. Mais une amélioration à cet effet a été proposée dans [SS03] en introduisant à chaque instant un nœud sans mesure en prenant compte des occultations ou les mauvaises détections dans un intervalle de temps inférieur à un seuil fixé.

2.2.2.2. Techniques probabilistes bayésiennes ou non de suivi d'objets

De nombreuses méthodes statistiques «*Kalman Filter*» et probabilistes (PDAF et MHT) ont été proposées dans l'état de l'art en réponse au problème de suivi d'objets mobiles. Le filtre de *Kalman* et ses extensions ont connu un grand succès pour leur exploitation dans plusieurs applications de suivi de cible pour leur élégance et robustesse (capacité prédictive et d'estimation). Le filtre de *Kalman* est basé sur deux

modèles qui sont le modèle d'observation (mesures du capteur) et le modèle d'état (modèle mathématique d'estimation) et très bien adapté lorsque le bruit dans l'image est important. Par exemple, dans le cas d'un objet de petite taille ou en présence d'images de types infrarouges, ce type de filtre a la capacité de filtrer des artefacts de faible durée comme les occultations ponctuelles. Ce filtre statistique prédictif a été utilisé avec un modèle d'apparence dans [RS98, PC01b] pour un suivi robuste d'objets mobiles.

En vision par ordinateur, le suivi d'objets dans une séquence d'images vidéo reste en investigation et en constante amélioration car il n'existe pas pour l'instant une technique généraliste s'adaptant à toutes les applications de vidéosurveillance dans un contexte réel [PC01, LWT03]. D'autres méthodes statistiques et probabilistes de suivi temporel multipistes qui exploitent plusieurs primitives simultanément existent. Parmi elles, nous pouvons citer:

- l'algorithme des plus proches voisins *NN* (*Nearest Neighbors*) qui met en correspondance les observations en fonction d'une distance;
- le filtre *Probabilistic Data Association Filter* (*PDAF*) estime le maintien de pistes [BSL95];
- le filtre *Joint Probabilistic Data Association Filter* (*JPDAF*) est une amélioration du précédent [Bla86, BSF88]. Le JPDAF traite les associations image par image, et non sur la durée, comme le font les méthodes MHT [CH96, SL93] et PMHT [SL94, Gel98];
- le filtre de *Kalman* Distribué DKF [RDS93] combine divers filtres de *Kalman* en parallèle pour augmenter la robustesse. Enfin, remarquons l'existence d'un filtre hybride qui est une combinaison de MHT et JPDAF utilise le MHT pour l'initialisation et la terminaison de pistes, et le JPDAF pour le maintien des pistes;
- Enfin, les techniques basées sur l'optimisation combinatoire.

2.2.3. Bilan de détection et du suivi pour l'approche adoptée

D'un point expérimental, nous avons étudié deux grandes classes d'approches de détection et de suivi d'objets mobiles dans des séquences vidéo. Chaque classe d'approche, que ce soit pour la détection ou le suivi est composée de deux techniques

qui font partie des plus connues et utilisées. Le Tableau 2.1 dresse un bilan des différentes méthodes de détection et de suivi d'objets mobiles exploités afin de faire ressortir les avantages et les inconvénients de chaque technique.

Ce bilan a guidé notre choix de techniques et d'algorithmes pour la détection et le suivi d'objet mobile afin de répondre à des contraintes d'exploitation efficace de ces résultats produits par ces couches de traitement et par les autres couches hautes de traitements d'interprétation automatique de la vidéo. Nous nous sommes focalisés sur trois principaux critères pour choisir nos méthodes.

Le premier critère concerne la difficulté d'initialisation de l'algorithme car cela permet d'avoir une approche moins contraignante, flexible simple et efficace.

Le second critère est relatif au coût en temps de calcul (aspects temps réel) et la complexité de l'algorithme tant pour le cas de la détection que pour celui du suivi. On note que le résultat du suivi dépend principalement de la qualité du celui de la détection.

Le troisième critère concerne la robustesse et l'efficacité des résultats, ce qui permet de calculer le taux d'erreur moyen du suivi afin d'évaluer la qualité de la détection par l'estimation du taux de fausse ou de non détection.

Ces trois critères sont très importants pour les outils mathématiques que constituent les modèles de Markov cachés choisis pour les traitements haut niveau notamment l'analyse de la vidéo.

Approches de détection	Avantages	Inconvénients	Approches de suivi	Avantages	Inconvénients
Différences d'images consécutives	Initialisation souple Simplicité et facilité d'utilisation Complexité algorithmique faible	Ombres rejetées Détection incomplète Mouvement obligatoire	Mise en correspondance basée sur l'association de deux images successives (MHT)	Temps de calcul acceptable Gestion d'apparition et de disparition de cibles	Temps de calcul faible Pourcentage de convergence faible l'erreur de suivi moyenne important
Soustraction de fond par modélisation statistique	Complexité algorithmique très élevée Classification très ardue Robustesse et efficacité des résultats	Initialisation avec une scène statique Ombres non rejetées	Mise en correspondance basée sur une statistique d'association(PDAF)	Temps de calcul acceptable Gestion d'apparition et de disparition de cibles	Temps de calcul très élevé Pourcentage de convergence élevé Erreur de suivi moyenne non négligeable
Soustraction de fond par moyenne récursive	Complexité algorithmique moyenne sinon faible Classification simple Robustesse et efficacité des résultats	Initialisation avec une scène statique Ombres non rejetées	Mise en correspondance basée sur une soustraction de fond par moyenne récursive	Temps de calcul acceptable Gestion d'apparition et de disparition de cibles	Temps de calcul moyen ou faible Pourcentage de convergence élevé Erreur de suivi moyenne minimale
Flot optique	Informations précises sur le mouvement Suivi robuste Possibilité de suivi avec prédiction	Complexité et calculs élevés, ombres non rejetées mouvement obligatoire Interprétation difficile	Mise en correspondance basée sur le mouvement calculé avec le flot optique	Temps de calcul acceptable Gestion d'apparition et de disparition de cibles	Temps de calcul élevé Pourcentage de convergence élevé Erreur de suivi moyenne minimale

Tableau 2.1. Tableau récapitulatif des méthodes de détection de mouvement et de suivi d'objets mobiles dans des séquences vidéo.

2.3. Les traitements haut niveau

Le module de traitements haut niveau exploite les résultats en sortie du module de traitement bas niveau combinés à des caractéristiques haut niveau qui sont extraites de la scène. Ces caractéristiques concernent les paramètres d'entrées des algorithmes d'analyse et d'interprétation de la scène qui produisent à leur tour en sortie des décisions de reconnaissance d'un événement puis d'une activité. Depuis les années 90, de nombreux algorithmes de reconnaissance d'activités humaines (scénarios intéressants) ont été mis au point en vision cognitive en proposition avec des applications diverses et dans plusieurs domaines [VVA05, ALD11].

Dans cette section, nous examinerons quelques travaux de recherche connexes sur la thématique de la reconnaissance d'activité humaine (*HAR : Human Activities Recognition*) en faisant un état de l'art sur les principales approches exploitées tout en mettant en exergue leurs avantages et leurs limites [Zia10]. Nous pouvons les classer en deux grandes catégories d'approches complémentaires et selon la communauté de chercheurs, nous distinguons deux principales classes d'approches qui sont utilisées pour reconnaître des scénarios spatio-temporels dans une séquence vidéo [ZMN08].

Nous avons les approches dites méthodes symboliques basées sur des techniques du domaine de l'intelligence artificielle et les approches dites méthodes orientées numériques qui sont basées sur des outils statistiques et probabilistes.

2.3.1. Approches orientées méthodes symboliques

Les approches orientées méthodes dites symboliques utilisent les termes scénario et chronique pour décrire les séquences. Ces techniques génèrent un processus de construction d'hypothèse et de vérification à partir de modèle d'enchaînement d'événements élaboré par les modules de traitements bas niveaux en vue de les décrire à un niveau décisionnel. La flexibilité et leur capacité d'adaptation aux contraintes des traitements de haut niveau pour la reconnaissance ont fait leur succès [Tes03, VBT03, SZP⁺11]. Cependant, ces techniques sont assez complexes à mettre en œuvre lorsque les contraintes temporelles doivent être intégrées dans le système de reconnaissance. De plus, la gestion des incertitudes et des imprécisions est quasiment impossible. Un état

de l'art assez complet de ces techniques a été proposé dans la thèse de T. Vu [Vu04] et l'Habilitation à Diriger des Recherches de Bremond [Bre07].

Dans la suite de la section, nous présentons les méthodes symboliques sans contrainte temporelle puis celles intégrant une contrainte temporelle et enfin une analyse de ces méthodes pour faire ressortir les avantages et les limites de ces méthodes.

2.3.1.1. Les méthodes symboliques sans contrainte temporelle

Les approches orientées méthodes symboliques sans contrainte temporelle élaborent un processus de raisonnement très proche du raisonnement humain. Mais loin de l'être, elles sont basées sur une représentation de graphes symboliques. Elles reposent sur des langages formels exploitant des règles pour transcrire les connaissances expertes d'un domaine. Ces langages abstraits permettent de gérer des contraintes logiques et par extension des contraintes temporelles entre événements et sont très proches des langages objets utilisés en programmation. Les modèles d'enchaînement d'événements (scénario) sont construits à partir de connaissances expertes à un haut niveau sémantique et respectent un raisonnement humain. Deux grandes catégories de ces techniques existent:

- les méthodes symboliques sans contrainte temporelle ;
- les méthodes symboliques avec contrainte temporelle.

Plusieurs travaux de recherche à travers des projets ont été réalisés dans le cadre de la reconnaissance de scénario en utilisant les méthodes symboliques. Parmi ces travaux, nous avons ceux de l'équipe Orion (Pulsar INRIA) [Bre97], et [TR99] qui proposent une approche basée sur un graphe d'états composé de quatre types de nœuds : objet, descripteur, opérateur et classificateur (voir Figure 2.3). Les activités des objets mobiles présents dans la scène sont mises en correspondance avec des scénarios composés de listes d'activités avec des contraintes logiques et temporelles. Un scénario reconnu est validé une fois que la séquence des activités de ce scénario est observée. La Figure 2.3

illustre le cas d'un scénario décrivant la rencontre de deux personnes près d'une cafetière. Le Tableau 2.2 décrit les nœuds et leur représentation dans le graphe d'état.

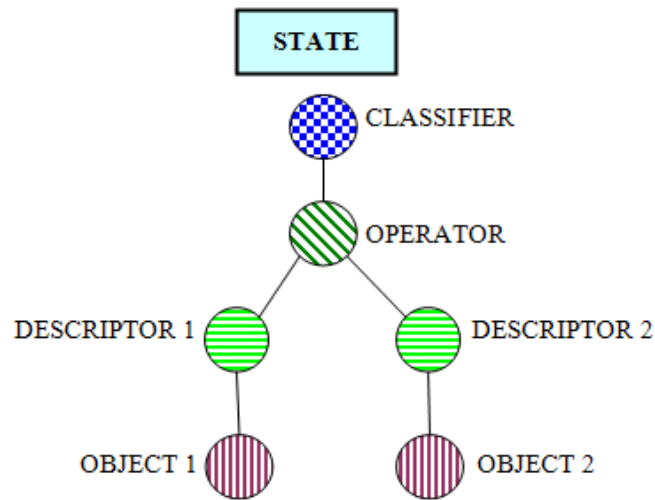


Figure 2.3. Modélisation du graphe d'état [TR99].

Noms des nœuds	Représentation des nœuds	Description
L'objet		Une observation dans l'environnement à un instant t
Le descripteur		Un vecteur d'observations telles que la taille, la position, la forme, la trajectoire et l'orientation.
L'opérateur		Un outil de combinaison de mesures de types arithmétiques ou logiques.
Le classificateur		Un outil de transformation numériques-symbolique

Tableau 2.2. Illustration de modélisation du graphe d'état [TR99].

L'une des premières approches basée sur un réseau de graphe symbolique dans lequel les nœuds correspondent à une variable booléenne permettant de reconnaître un scénario a été proposée par [KAA86], et offre une technique de reconnaissance avec détermination des événements [KS86].

Cependant, la technique est difficile à exploiter pour reconnaître des actions dans les applications du monde réel. D'autres approches basées sur les réseaux de Pétri ont

été développées par [CCT96]. Ces approches ont connu quelques succès en raison de la capacité des réseaux de Pétri à modéliser des systèmes à événements discrets. Le projet de perception a mis en place un formalisme des réseaux de Pétri interprétés pour décrire à partir de modèle d'enchaînements d'activités la couche d'interprétation symbolique de la scène à travers un processus de génération-vérification [CCT96].

Ghanem et al. [NGD04] ont développé une approche basée sur la définition d'événements primitifs qui déterminent les événements d'intérêts en y intégrant tant bien des contraintes logiques que temporelles. La phase de détection et de suivi de la cible exploite des événements primitifs qui sont représentés à l'aide d'un réseau de Pétri. Une approche basée sur un seul réseau de Pétri a été utilisée pour reconnaître un seul type d'activité mais incapable de reconnaître d'autres occurrences de la même activité.

Par exemple, dans ce contexte, Lesire et Tessier [LCT07] ont conçu un réseau de Pétri pour reconnaître une activité donnée dans laquelle les nœuds du réseau correspondent à des événements typiques et les jetons aux objets mobiles impliqués dans la réalisation de l'activité. Une grammaire stochastique a été proposée par Ivanov et Bobick en 2000 [IB00] pour analyser des actions simples afin de les reconnaître par des modules de vision haut-niveau.

2.3.1.2. Les méthodes symboliques avec contraintes temporelles

Les techniques symboliques [SAA12] avec contraintes temporelles permettent de modéliser et de reconnaître des scénarios intégrant des contraintes temporelles. Ces techniques sont basées sur différentes modélisations temporelles et exploitent des raisonnements distincts [VBT02]. Dans ces approches, les modélisations peuvent varier selon qu'elles autorisent ou non et à différents niveaux, la représentation de l'incertitude et du temps. Les modélisations sont généralement classées selon les relations temporelles qui les lient et peuvent être numériques, symboliques ou un mélange gaussien. Il existe plusieurs types de raisonnements qui exploitent deux principales approches telles que les approches de type CSP (*Constraints Satisfaction Problem*) et les approches graphiques.

Dans le cadre de l'exploitation des réseaux de Pétri, deux extensions des réseaux de Pétri ont été développées en réponse à l'intégration de la contrainte temporelle. En outre, une autre approche basée sur des techniques de représentation du temps par intervalle de temps. Nous avons les réseaux de Pétri temporisés basés sur la durée d'état et les réseaux de Pétri temporels basés sur l'estimation du délai entre deux événements.

De nombreux modèles basés sur les Chroniques ont été développés parmi lesquels nous avons l'approche proposée par C. Dousson [Dou94] et M. Ghallab [CDG93] qui est un système de reconnaissance de chroniques basées sur le langage IxTet et appliquées à des problèmes de supervision dans les réseaux de télécommunications. Cette technique appelée *CRS (Chronicle Recognition System)* a été améliorée en y intégrant une architecture hiérarchique par [DM07] en cas de description plus complexe du système. Deux catégories de techniques réseaux symboliques sont basées sur les chroniques que sont la méthode *SPRS (Stores Partially Recognized Scenarios)* qui est une approche de reconnaissance de scénario basé sur une analyse des scénarios reconnus par le système par le passé [VTB02, TR00], et la méthode *STRS (Stores Totally Recognized Scenarios)* qui est une approche de reconnaissance de scénario basée sur une analyse des scénarios reconnus par le système ou qui peuvent être reconnus à l'avenir [Gha96].

La méthode *STRS* permet de reconnaître un scénario en cherchant dans l'ensemble des scénarios précédemment comptabilisés et les sous-scénarios correspondant au modèle de scénario pour être reconnu. Ainsi, si le système ne reconnaît pas le scénario, il faudra recommencer le même processus (révérifier les mêmes contraintes) dans l'instant suivant, ce qui implique un temps de traitement coûteux. Cette principale conséquence est le premier problème auquel on fait face dans l'utilisation de la méthode *STRS*. Un deuxième problème est que les algorithmes *STRS* ne peuvent stocker et maintenir que toutes les occurrences de scénarios précédemment reconnus. La méthode *SPRS*, à partir d'un scénario reconnu permet de prédire les scénarios qui devraient être reconnus dans les prochains instants. Ainsi, les scénarios doivent être délimités dans le temps pour éviter les scénarios attendus sans fin.

Un deuxième problème est que les algorithmes SPRS doivent stocker et maintenir toutes les occurrences de scénarios partiellement reconnus, ce qui implique un espace de traitement coûteux. Par exemple, en 2000, N. Rota et M. Thonnat [RT00] ont utilisé une représentation déclarative de scénarios définis comme un ensemble de contraintes spatio-temporelles et logiques. Ils ont utilisé une technique traditionnelle de résolution de contraintes pour reconnaître des scénarios. Pour réduire le temps de traitement de l'étape de reconnaissance, ils ont proposé une technique de vérification de la cohérence du réseau sous contraintes en utilisant l'algorithme AC4 [MH86].

En 2002, R. Gerber, H. Nagel et H. Schreiber [GNS02] ont défini une méthode pour reconnaître un scénario basé sur une logique temporelle floue. Dans la même année, T. Vu, F. Bremond et M. Thonnat [VBT02] présentent une approche pour optimiser la résolution de la contrainte temporelle, en ordonnant dans le temps les sous scénarios du scénario pour être reconnus. La caractéristique commune de ces approches est de stocker tous les scénarios totalement reconnus mais par le passé. Il existe d'autres approches symboliques qui sont associées aux langues avec des ontologies de représentation proposées par Médioni [BM98, HTWM04], Hongeng et Nevatia [Hon02, HNB04, NHB04].

2.3.1.3. Limites des méthodes symboliques

La difficulté majeure des méthodes symboliques est qu'elles ne gèrent ni l'incertitude, ni l'imprécision des données (observations) et celles générées par les traitements du processus de vision par ordinateur. De plus, il est très difficile, voir impossible de prendre en compte le caractère incomplet de la connaissance. Dans le cadre de l'utilisation de ces méthodes, chaque état dans le graphe est associé à un événement primaire et représente une variable binaire. Ce dernier indique que l'événement associé à l'état-graphique apparaît ou n'apparaît pas. Par conséquent, ces méthodes ne fournissent pas la solution au processus sémantique faible niveau en charge des mécanismes de vision par ordinateur pour la détection d'événements qui sont très importants quand il s'agit de faire des ajustements, comme une rétroaction sur l'ensemble du système [Kun04].

En plus de ces nombreuses difficultés, pour la gestion de multiples contraintes, les méthodes symboliques peuvent facilement conduire à des explosions combinatoires dans le développement d'algorithmes avec une complexité de calcul énorme. Toutefois, en fonction de l'application visée, ces approches sont très appropriées, efficaces et ont montré d'excellents résultats. En conséquence, l'incapacité de gestion des incertitudes de ces techniques ne permet pas de répondre à certaines exigences du contexte applicatif de nos travaux de thèse. Pour répondre à ces exigences, d'autres approches orientées méthodes numériques (déterministe ou statistique) ont été proposées afin de gérer efficacement, autant l'incertitude que l'imprécision des données traitées. Elles peuvent également être utilisées pour résoudre le problème récurrent de l'incomplétude de la connaissance qui est au cœur de l'expertise et des applications médicales.

L'exemple du contexte de notre application qu'est celui du domaine médical spécifiquement, la conception et la mise place d'un système d'assistance par ordinateur à la surveillance des patients aux USIs pour aider l'opérateur médico-surveillant (infirmière superviseur) l'illustre bien. Ce système à partir de génération d'alarme visuelle en ligne suite à la détection automatique d'un événement important ou anormal (situation à risque pour l'état du patient, dangereux pour la santé du patient, par exemple *la détection d'agitation ou de douleur*), permet d'archiver tous les scénarios reconnus comme importants dans une base de données spéciale nommée boîte noire médicale. Les méthodes numériques sont plus adaptées et nous ont permis d'avoir des résultats satisfaisants tout en gérant l'ensemble des imperfections liées aux données et celles générées par les traitements de bas niveau. Dans la section ci-dessous, nous faisons un bref état de l'art des méthodes orientées numériques.

2.3.2. Approches orientées méthodes numériques

De nombreuses approches orientées méthodes numériques exploitent des outils mathématiques statistiques et ont été proposées par la communauté de vision par ordinateur pour reconnaître les activités humaines dans des séquences vidéos [BPV04, HBN00, RRL89, HJB98, TBM⁺04]. Un des plus grands défis de ces techniques est de relier les observations numériques aux états du système observé en tenant compte de l'incertitude sur les observations et les états. A l'opposé des méthodes symboliques,

celles numériques permettent de gérer l'incertitude sur les observations et les états. Les méthodes numériques axées sur l'analyse de séquences d'images et l'interprétation de la scène vidéo peuvent être classées en deux catégories: les approches déterministes et les approches probabilistes. Cependant, il convient de noter que cette dernière catégorie est la plus utilisée ou exploitée par les chercheurs pour la reconnaissance de scénario. Dans les approches déterministes comme la programmation dynamique. Elle fournit en particulier une technique efficace pour associer deux séquences de primitives ordonnées globalement. En reconnaissance de la parole, cette technique déterministe a été utilisée puis adaptée en vision par ordinateur afin de mettre en correspondance des primitives géométriquement alignées. La technique a été également adaptée dans le cadre de la stéréovision afin de mettre en correspondance des primitifs se situant sur les droites épipolaires [Zia10].

En vision par ordinateur, une des approches naturelles pour la reconnaissance des activités humaines consiste à utiliser des outils probabilistes [BPV04, HBN00, Rab89]; ou les réseaux de neurones [HJB98] afin de faire face à la gestion des incertitudes propagées via les processus de traitement de vision bas niveau [ZM11] et l'imprécision des observations. Les nœuds du réseau correspondent généralement à des scénarios qui sont reconnus à un instant donné avec une probabilité calculée. Ces techniques statistiques sont basées sur l'estimation du modèle de séquence d'événements qui représentent le scénario appris a priori lors de la phase d'apprentissage. A partir des données d'apprentissage, généralement, des observations numériques peuvent être étiquetées dans le cas d'un apprentissage supervisé ou non étiquetées dans le cas de l'apprentissage non supervisé [ZM11].

À cette fin, ces techniques ont besoin en amont, d'une étape qui concerne généralement l'extraction de caractéristiques temporelles et/ou fréquentielles des données. Ces caractéristiques sont ensuite utilisées pour classer les activités, en se basant exclusivement sur des méthodes d'apprentissage automatique supervisées. Parmi les nombreuses méthodes d'apprentissage utilisées pour ces applications, on peut citer: le k-plus proche voisin [ABT10], les modèles de Markov cachés (*Hidden Markov Models : HMM*) [HLT07]; les réseaux de neurones artificiels (*Artificial Neural Network : ANN*) [HJB98] les machines à vecteurs de support (*Support Vector Machine : SVM*)

[ABT10, YCB11]. En outre, la phase d'apprentissage permet également d'estimer les liens statistiques entre états à un même instant et à des instants consécutifs. Elle permet par conséquent d'élaborer des modèles d'évolution du scénario à reconnaître. Ce type d'approche a été utilisé au départ en reconnaissance de la parole [Rab89]. Cependant, une caractéristique importante des méthodes d'apprentissage automatiques supervisées est que la représentation des activités n'est possible que pour un certain nombre d'activités prédéfinies qui ont déjà été étudiées et étiquetées. Ce qui rend ces procédés difficiles à modifier ou à mettre à jour. L'ajout d'une nouvelle catégorie d'activités, par exemple, nécessite un nouvel ensemble de données et/ou le recyclage de l'ensemble du système (un apprentissage total de la base de reconnaissance afin d'intégrer les nouveaux scénarios). Une autre caractéristique de la plupart des méthodes d'apprentissage automatique supervisé standard dans un contexte probabiliste est qu'elles conduisent à une «boîte noire» à l'intérieur de laquelle les décisions sont prises de manière invisible à l'utilisateur. Ce qui est particulièrement gênant pour les applications dans le domaine médical, où la compréhension du raisonnement derrière la décision rendue est extrêmement importante pour sa validation [SAA12].

Dans les travaux d'état de l'art, de nombreux chercheurs se sont intéressés à la détection et l'analyse du comportement humain dans des séquences vidéo en utilisant des outils probabilistes et statistiques. Parmi les outils probabilistes stochastiques ou simplement probabilistes, les plus couramment exploités sont les techniques basées sur le modèle de mélange gaussien (*Gaussian Mixture Model : GMM*) pour la décision, les réseaux de neurones artificiels, le réseau bayésien dynamique (DBN), les modèles de Markov cachés avec plusieurs extensions dont l'une des dernières en date est celle crédibiliste qui exploite les fonctions de croyances transférables.

Dans [NCM04] en 2004, Mckenna et Nait-Charif ont développé un système de détection de «*chute de personne*» à l'intérieur d'une pièce d'un appartement en utilisant les modèles de mélange gaussien (*Gaussian Mixture Model : GMM*) pour représenter les zones d'intérêts en associant l'algorithme de maximisation (*Expect Maximisation: EM*) pour l'étape d'estimation et un estimateur de maximum a posteriori (MAP) pour la détermination de la structure et les paramètres du modèle. Cette technique permet de combiner l'information a priori sur les zones de détection et les composantes

gaussiennes. Dans les zones d'intérêts, ils ont défini des activités telles que «*Entrez par la porte de la salle*», «*S'asseoir sur le sofa*» et «*Sortir par la porte arrière*». La Figure 2.4 est une illustration de la détection du scénario «*chute de personne*» avec une segmentation de la trajectoire.

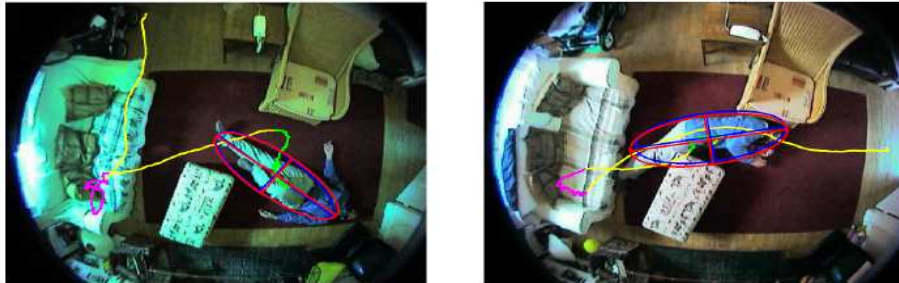


Figure 2.4. Détection d'une chute avec trajectoire segmentée [NCM04].

2.3.2.1. Les réseaux de neurones

L'utilisation des réseaux de neurones pour la reconnaissance des activités humaines, remonte aux années 90. Les réseaux de neurones appelés aussi réseaux connexionnistes sont souvent utilisés pour les problèmes de classification. Par exemple, Howell et Buxton [HJB98] d'une part, et d'autre part, Howarth et Buxton [HJB00] ont utilisé les réseaux de neurones pour la reconnaissance de geste humain [WB01] dans un contexte d'interaction visuelle médiatisée. Une approche consiste à modéliser un comportement humain par une séquence temporelle de mouvements du corps ou les configurations du corps humain tel qu'un changement d'une pose de tête, et des actions comme «*marcher*», et «*s'asseoir*» [VT04]. Plusieurs exemples de mouvements et comportements du corps humain sont illustrés dans le Tableau 2.3 pour la détection de gestes [WB01] pour la communication humaine. Quelques années plus tard, les mêmes auteurs proposent d'utiliser une autre variante de réseaux de neurones basée sur les fonctions de base radiale (*Radial basis Functions : RBF*) et ont proposé par la suite une version améliorée de la RBF qui intègre la gestion des contraintes temporelles appelées fonctions radiales temporisées (*Time-Delay Radial Basis Functions: TDRBF*) pour une description plus fine du scénario à reconnaître [HJB02]. Le TDRBF a été utilisé pour reconnaître de simples gestes de pointage et des gestes d'indication communicationnelle de la main dans des séquences d'images. Les réseaux de neurones peuvent stocker une base de connaissances résultant de l'analyse de l'environnement

formel où se déroule la scène. Les réseaux de neurones ont la capacité de stocker une base de connaissances résultant de l'analyse de l'environnement formel où la scène se déroule et permet d'éviter par conséquent l'étape ardue de formalisation du modèle de scénario. D'autres avantages de ces outils résident dans l'utilisation conjointe de plusieurs neurones pour la représentation de connaissances à la phase d'apprentissage. De plus, il est possible qu'un neurone participe pour le codage de plusieurs d'entre eux. D'autres types d'extensions comme les réseaux de neurones flous ont été exploités en reconnaissance de la parole et ont permis de combler plusieurs limites du modèle des réseaux neurones standard [EEG01].

Cependant, les réseaux de neurones artificielles ne sont pas efficaces pour faire face à des problèmes de reconnaissance de comportements complexes qui font intervenir un grand nombre d'objets physiques avec des contraintes temporelles complexes (par exemple, lorsque les contraintes temporelles sont synchronisées). Ces outils peuvent conduire facilement à une explosion combinatoire des comportements possibles correspondant à toutes les combinaisons probables des objets physiques détectés dans la scène [VT04].

Gesture	Body Movement	Behavior
<i>pntrl</i>	point right hand to left	pointing left
<i>pntrr</i>	point right hand to right	pointing right
<i>wavea</i>	wave right hand above head	urgent wave
<i>waveb</i>	wave right hand below head	non-urgent wave

Tableau 2.3. Illustration de quelques exemples de mouvement du corps et définitions de comportement pour la base de données de geste [HJB02].

2.3.2.2. Les réseaux bayésiens dynamiques

Un réseau bayésien dynamique (*Dynamics Bayesian Network: DBN*) est composé de graphes et de sous-graphes temporellement séparés et reliés entre eux par des liens dynamiques [ORP00]. Ces sous-graphes sont reconfigurables automatiquement et comprennent un réseau de nœuds qui peuvent être reliés non seulement aux autres nœuds de la même tranche de temps, mais aussi à des nœuds de la tranche d'intervalle de temps suivante ou précédente. D'autre part, il est aussi possible de définir un réseau

bayésien dynamique comme une extension des réseaux bayésiens. Ces derniers ont été exploités pour la première fois pour la reconnaissance de comportements dans une application de surveillance à distance. Ils fournissent un formalisme cohérent et intuitif pour représenter l'incertitude et l'imprécision des informations. La capacité des réseaux bayésiens à représenter les données d'un problème à la fois sous une forme qualitative et quantitative pour les systèmes de reconnaissance des activités humaines justifie le grand intérêt témoigné pour eux dans leur exploitation par les chercheurs en vision par ordinateur ces dernières années. Chaque type de représentation qualitative et quantitative désigne les deux principaux niveaux de représentation des réseaux bayésiens. Les mécanismes de raisonnement dans les réseaux bayésiens dynamiques peuvent être regroupés en deux niveaux d'abstractions :

- 1) le niveau qualitatif est constitué d'un graphe acyclique orienté dans lequel les nœuds représentent les variables qui peuvent prendre plusieurs valeurs, et des arcs des relations causales entre les variables ;
- 2) le niveau quantitatif concerne la table de probabilités associées à chaque nœud qui définit l'état du nœud étant donné l'état de ses parents. La relation de cause à effet entre les variables est exprimée par la présence ou l'absence d'arcs entre eux.

Un DBN «*Dynamic Bayesian Network*» est représenté par un graphe acyclique fini, illustré comme exemple à la Figure 2.5 et défini par :

- Des nœuds qui sont les variables aléatoires discrètes;
- Des arcs qui permettent de relier graphiquement les nœuds;
- Des bords qui sont des influences directes sur les nœuds;
- Un tableau de distribution de probabilité est associé à chaque nœud représentant une distribution de probabilité conditionnelle (*Conditional Probability Distribution : DPC*) pour quantifier l'effet que les nœuds parents ont sur les nœuds fils. Le CPD est estimé par l'équation 2.5 :

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i / pa(X_i)) \quad (2.5)$$

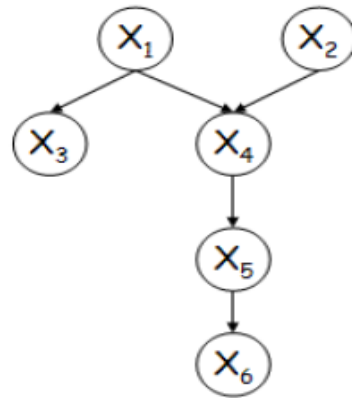


Figure 2.5. Exemple de représentation d'un réseau bayésien dynamique [Zia10].

Les réseaux bayésiens sont principalement utilisés pour calculer les probabilités conditionnelles de variables liées les unes aux autres par des relations de cause à effet. Cette utilisation est appelée inférence. Un type d'inférence particulier est la mise à jour des probabilités qui se produisent en raison d'une série d'observations. L'inférence revient à un problème de calcul déductible et parfois de façon incrémentale avec une solution mathématique existante en général. Toutefois, selon la structure du réseau, ces calculs sont plus ou moins complexes.

Par exemple, en 1995, le travail de Buxton et Gong [BG95] repose sur l'utilisation des réseaux bayésiens pour caractériser les positions et les interactions entre les objets observés dans une zone de surveillance de trafic. Trois années plus tard en 1998, Remagnino et al. ont utilisé les réseaux bayésiens pour la surveillance de parking [TRB98]. De même, Hongeng et al. [HN01], [SHB03b] ont également exploité les réseaux bayésiens combinés à une architecture multi-agent dans la reconnaissance des événements complexes. Ces événements sont représentés avec une couche hiérarchique. De même en 2005, Peursum et al. [WPV05] ont modélisé les interactions entre les personnes et les objets détectés dans une scène en utilisant les réseaux bayésiens.

2.3.2.3. Les chaînes de Markov cachées et extensions

Les modèles de Markov cachés (*Hidden Markov Model: HMM*) représentent l'un des outils les plus puissants pour la modélisation, l'analyse et l'exploration de données séquentielles. Ils représentent l'une des approches probabilistes les plus utilisées ces

trois dernières décennies pour la reconnaissance des activités humaines en vision par ordinateur. Plus récemment, les modèles probabilistes graphiques dynamiques [Mur02] tels que les réseaux bayésiens et ceux dynamiques sont apparus avec une très bonne performance en terme de la modélisation, de l'apprentissage des scénarios et du calcul de l'inférence.

Les HMM sont très appropriés pour modéliser les états d'un système dynamique dans le cas où ces états ne sont pas directement observables. Deux principaux types d'HMM existent notamment : le type continu et le type discret. Une caractéristique intéressante des HMM est leur invariance à l'échelle de temps permettant ainsi de reconnaître les activités avec différentes échelles de temps. De nombreux problèmes d'applications complexes ont été résolus avec succès en utilisant les HMM. Nous pouvons noter par exemple, en reconnaissance de la parole, la découverte de séquence de mots dans les enregistrements de la parole [Rab89], en reconnaissance d'activités humaines dans des vidéos, la reconnaissance de séquence de mouvements humains. De nombreuses variantes des modèles de Markov cachés ont été développées en réponse aux lacunes identifiées dans le modèle classique en fonction des exigences des applications.

Le type discret du modèle de Markov caché a été exploité par exemple en bioinformatique pour la reconnaissance des séquences de bases azotées sur les brins d'ADN. Il existe quatre observations possibles discrètes qui sont A, C, G et T. Les quatre bases azotées s'alternent pour décrire les séquences d'ADN présents dans chaque brin et établissant ainsi un message porteur d'une information génétique. Dans ce modèle, l'HMM est utilisé pour aligner une séquence inconnue de bases azotées avec d'autres connues pour déterminer le rôle et les fonctions des séquences analysées. Ce mécanisme permet de détecter de nouvelles molécules ou de décoder des génomes et a notamment été mis en œuvre sur des architectures à base de circuits électroniques reconfigurables pour gérer de grandes quantités de données [OSJM07].

Dans le cadre de la reconnaissance de l'activité humaine, une première application a été réalisée via un système de reconnaissance de gestes basée sur des HMM et dans ce même contexte d'application, Ymato et al. [OYI92] ont proposé une approche de reconnaissance de six gestes de tennis par trois sujets différents en utilisant

un modèle discret d'HMM. Pour le tennis, les événements prédéfinis et prédéterminés à reconnaître dans les vidéos sont : les coups joués (revers, coup droit, smash) et les points marqués (ace, monté au filet). Les HMM ont permis de développer plusieurs applications de reconnaissance d'actions dans des vidéos de sport [KRD⁺06], [DVB⁺05].

De même, en 2004, Bui et al. ont proposé un modèle de Markov caché hiérarchique (*Hierarchical Hidden Markov Models : HHMM*) étendu à un modèle hiérarchique général toujours pour la reconnaissance automatique de scénario dans une scène vidéo. Une autre approche a été développée par Bengio et Frasconi [BF95], où ils proposent un modèle de Markov caché avec entrée/sortie (où IOHMM est mis pour *Input/Output HMM*). L'IOHMM est une extension du modèle de Markov caché classique et a été utilisé pour aligner une séquence d'entrée à une séquence d'états discrets afin de générer la séquence de sortie. Inanov et Bobick [IB00] ont traité séparément la reconnaissance des scénarios sur une partie et les actions primitives en deuxième partie. Ils ont utilisé des HMM pour la reconnaissance des scénarios comme par exemple «voiture entrant, arrêt de voiture, personne sortante» et une grammaire spécifique pour la reconnaissance scénario (comportement complexe).

D'autres extensions du modèle HMM de base ont également été utilisées. Parmi elles, nous avons, les modèles de Markov cachés couplés (*Couple Hidden Markov Model : CHMMs*) pour modéliser les interactions humaines (Oliver et al., 2000), et les modèles de Markov de longueur localement variable (*Variable Length Markov Models : VLMMs*) pour optimiser la taille des modèles de comportement [GJH01]. Récemment, d'autres études ont porté sur la modélisation des trajectoires des objets observés. Par exemple, F. Porikli [Pro04] a développé une approche de classification basée sur les HMM pour modéliser le lien de causalité temporelle entre les trajectoires traitées. Cette modélisation prend en compte les orientations locales successives des trajectoires afin de les comparer en termes de la forme et de la vitesse. Un mélange de fonctions gaussiennes a été rajouté afin de sélectionner un certain nombre d'états en tenant compte de la taille des trajectoires. Dans cette application, un atout majeur des HMM a été l'intégration de l'aspect dynamique des transitions des trajectoires, qui s'applique bien à la reconnaissance.

En 2005, Duong et al. ont proposé un modèle semi-Markovien caché commuté pour la reconnaissance d'activités et la détection de situations anormales [DVB+05]. Les mêmes auteurs en 2009, ont proposé une modélisation de la durée dans un modèle d'HMM hiérarchique pour reconnaître les activités humaines [DVB+09].

D'autres outils de la théorie de l'incertain tels que la théorie de l'évidence et le modèle de croyances transférables, ont permis de mettre en place de nouvelles extensions des HMM comme les modèles de Markov Cachés Évidentiels (*Evidential Hidden Markov Modèle : EvHMM*). Une application de ces extensions a été proposée par E. Ramasso en 2007, dans le cadre de l'analyse et l'interprétation des vidéos d'athlétisme pour la reconnaissance de type de saut (en longueur, en hauteur puis triple saut par exemple).

2.3.3. Bilan des techniques de reconnaissance de scénario

La revue bibliographique des outils et techniques pour la reconnaissance de scénarios dans des séquences d'images vidéos a permis de distinguer deux grandes catégories d'approches obtenues selon la communauté de chercheurs qui les exploite. Une première catégorie dite méthodes symboliques sont généralement des approches de type qualitatif et ont permis d'avoir des résultats de reconnaissance très encourageants. Cependant, ils peuvent conduire facilement une explosion conditionnelle en présence de scénarios complexes ou lorsqu'il y a un nombre élevé d'objets dans la scène. L'inconvénient majeur de cette catégorie de techniques est la difficulté ou l'impossibilité de gestion d'incertitude.

En outre, la seconde catégorie qui concerne les méthodes numériques représente des techniques quantitatives car ils exploitent des outils mathématiques de type probabiliste et statistique. Ces techniques sont très efficaces et robustes et ont permis d'avoir d'excellents résultats en reconnaissance d'activité. De plus, en fonction du cadre de raisonnement, ils offrent une flexibilité très intéressante dans la gestion d'incertitude et d'imprécision et d'incomplétude. Le problème majeur de ces outils est relatif à la phase d'apprentissage fastidieuse, la difficulté de mise à jour de la base de connaissance et enfin un coût de temps de calcul trop élevé. Le Tableau 2.4 dresse un bilan comparatif

des différentes techniques de reconnaissance de scène en ressortant les avantages et les inconvénients de chaque technique.

	Méthodes	Avantages	Inconvénients
Méthodes symboliques			
1	Classification d'action	Simplicité algorithmique Facile à utiliser	Explosion conditionnelle pour des scénarios complexes
2	Les automates	Peut traiter efficacement des séquences d'événements	Explosion conditionnelle en fonction du nombre d'objets présents dans la scène Impossible de représenter des événements synchronisés
3	Approche CSP	Simplicité algorithmique Facile à utiliser	Complexité algorithmique prohibitive
Méthodes symboliques et temporelles			
5	Réseau de Pétri	Permet une prédiction Peut traiter efficacement des séquences d'événements	Ne peut pas traiter les événements interdits Explosion combinatoire en cas de problème d'objet multi-physique
6	Reconnaissance de chronique	Très efficace pour le problème mono-physique-objet Efficace pour le traitement des événements négatifs	Explosion combinatoire en cas de multi-physique-objet problème
7	Approche TCP	Peuvent utiliser efficacement les résultats calculés dans le passé Permet la prédiction	Explosion combinatoire du nombre d'objets physiques
Méthodes Orientées Numériques			
8	Classificateur Bayésien	Gestion d'incertitudes	Difficultés de gestion des contraintes temporelles et coût de calcul élevé
9	Réseau de neurones	Adaptation à l'environnement Gestion de données incomplètes	Difficultés de gestion des contraintes temporelles et Coût de calcul élevé
10	MMCs	Gestion d'incertitudes Reconnaissance de séquences d'événements	Phase d'apprentissage fastidieuse Ne peuvent pas faire face à des relations temporelles complexes Coût de calcul élevé
Méthodes orientées numériques à contraintes temporelles			
11	MMCs Hiérarchiques	Gestion d'incertitudes Reconnaissance de séquences d'événements Gestion de la structure hiérarchique des actions	Coût de calcul élevé Complexité algorithmique élevé
12	MMCs temporel	Gestion d'incertitudes Reconnaissance de séquences d'événements Gestion des relations temporelles complexes	Coût de calcul élevé Complexité algorithmique élevé
13	MMCs Crédibilistes	Gestion d'incertitudes et d'imprécisions Gestion d'incomplétude des données (modélisation de l'ignorance) Gestion du doute et du conflit Combinaison et fusion d'informations	Coût de calcul élevé Complexité algorithmique élevé

Tableau 2.4. Bilan comparatif des approches de reconnaissance automatique d'activité.

2.4. Module de gestion d'incertitude

Les connaissances manipulées par le système d'interprétation sont généralement imparfaites. Elles peuvent être imprécises, incertaines ou incomplètes. L'incertitude est relative à la vérité d'une information. En d'autres termes, elle exprime un doute sur la validité de cette information. L'imprécision est liée au contenu de l'information et représente un défaut quantitatif [Blo96]. L'incomplétude représente l'absence d'une partie de l'information.

L'objectif de la gestion de l'imperfection, englobée plus généralement dans le thème «*la gestion de l'incertitude*» est de générer, à partir de données imparfaites, des conclusions satisfaisantes ou nuancées [BMN96]. Cette gestion de l'incertitude permet au système de manipuler des décisions nuancées mais sûres, plutôt que des décisions fermes qui peuvent être erronées. Une modélisation des imperfections est obtenue généralement sous forme d'une quantification de l'incertitude ou de l'imprécision.

Dans un système d'interprétation, les imperfections sont multiples et proviennent des données brutes issues de la vision bas niveau et des différentes connaissances ou modèles exploités dans la chaîne d'interprétation. Une représentation explicite par exemple sous forme de mesures d'incertitudes ou d'imprécision permet de mieux contrôler la manipulation des informations. Divers cadres de gestion de l'incertitude sont exploitables en interprétation. Ils exploitent des outils mathématiques dont le cadre théorique est appelé «*les théories de l'incertain* » et sont très exploités dans le cadre d'exploration de données via des processus de fusion ou de combinaison d'information pour améliorer une prise de décision par la communauté de la «*Fusion d'Informations*» [DP87b]. On y retrouve la théorie des probabilités, la théorie des ensembles flous et des possibilités, et la théorie de l'évidence et son extension qu'est le modèle de croyances transférables.

Le choix du cadre théorique est très souvent motivé par le type d'information à traiter. Une information peut être considérée comme «*forte*» si elle est réellement mesurée à partir de fréquences d'occurrences d'événements. Par contre, une information de type subjectif est considérée comme «*faible*».

2.4.1. Cadre de raisonnement probabiliste

L'approche probabiliste, qui constitue l'approche normative de base, est considérée comme la plus exigeante. En effet, elle impose que de nombreuses lois soient connues et vérifient des contraintes strictes de cohérence. Un désavantage de la théorie classique des probabilités est qu'elle ne modélise pas les notions d'ignorance et d'imprécision de manière naturelle [Sha81]. En revanche, cette modélisation conduit aux solutions optimales lorsqu'elle est appliquée dans des conditions idéales de connaissance des phénomènes observés et de leurs perturbations (information forte).

La modélisation à partir de réseaux bayésiens, qui est issue des approches probabilistes, est considérée comme un compromis entre les approches normatives et descriptives. Cet outil propose un formalisme cohérent et intuitif pour représenter des informations incertaines et imprécises. L'attrait des réseaux bayésiens réside dans leurs capacités à représenter à la fois les données qualitatives et quantitatives [Pea88]. La topologie du graphe du réseau représente la nature qualitative de la relation de causalité entre les nœuds.

Les approches de la reconnaissance de scénario fondées sur la théorie probabiliste ont été suffisamment exploitées et fournissent des outils appropriés pour la gestion de l'incertitude et de l'imprécision. Néanmoins, les autres théories de l'incertain n'ont été suffisamment exploitées et ni assez bien évaluées sur le problème de la reconnaissance de scénario. La théorie des possibilités, couplée avec les ensembles flous, et la théorie de l'évidence avec son extension aux modèles de croyances transférables constituent aujourd'hui des cadres génériques et assez puissants à exploiter pour la gestion de l'incertitude.

2.4.2. Cadre de raisonnement possibiliste

La théorie des possibilités et les ensembles flous font partie des approches descriptives [Dub87a]. Ces approches modélisent de manière assez naturelle l'incertitude et l'imprécision. Cette modélisation peut être basée sur des informations subjectives mais aussi construites à partir de données statistiques. Comme la théorie de

l'évidence, elle a l'avantage de pouvoir représenter de manière explicite la notion d'ignorance en considérant que toutes les hypothèses sont possibles.

En revanche, pour la théorie des probabilités, la variable probabiliste est distribuée de manière équitable sur toutes les hypothèses ambiguës, ce qui peut introduire des incohérences dans le raisonnement. Comparée aux autres cadres, la théorie des possibilités présente une grande liberté quant aux choix des opérateurs de combinaisons. Le choix d'un opérateur, qui se fait généralement en fonction du comportement recherché, peut être considéré comme un avantage [Dub87a]. Mais il peut être aussi perçu comme un inconvénient car il impose des justifications qui sont souvent liées au contexte d'utilisation.

2.4.3. Cadre de raisonnement crédibiliste ou évidentiel

La théorie de l'évidence [Dem67, Shaf76], qui est considérée également comme une approche normative, permet de modéliser l'incertitude et l'imprécision. Un de ses principaux avantages est qu'elle permet de modéliser l'ignorance en distribuant des masses sur un ensemble d'hypothèses plutôt que sur des singletons comme l'approche probabiliste [Dem68]. Toutefois, elle reste rigide en ce qui concerne la combinaison des informations puisqu'il n'existe que peu de règles de combinaisons.

Récemment, un nouveau cadre de gestion de l'incertitude basé sur la théorie de l'évidence avec une extension qui concerne l'élaboration et la mise en place de nouveaux opérateurs prudents pour la phase de fusion d'informations et de nouvelles fonctions de prise de décision ont été proposées en exploitant les fonctions de croyances transférables. Les règles de combinaison qualifiées de rigides dans la théorie de l'évidence sont devenues très souples et permettent de faire une fine description des états du modèle. Ils modélisent le doute et le conflit entre les états du système, de prendre en compte l'ignorance. Ces mécanismes sont proposés dans un nouveau modèle de la théorie de l'évidence appelé le Modèle des Croyances Transférables (*TBM : Transferable Belief Model*) qui est un cadre formel générique développé par Ph. Smets [SK94] pour la représentation et la combinaison des connaissances.

Le TBM est basé sur les fonctions de croyance et propose un ensemble d'opérateurs permettant de combiner des fonctions de croyances fournies par des

sources d'informations pouvant être complémentaires, redondantes et éventuellement non-indépendantes.

Le TBM permet de représenter explicitement le doute sur les états, ce qui s'avère intéressant pour la modélisation explicite des transitions entre états, que ce soit lors de l'apprentissage de modèles ou lors du processus de reconnaissance.

De plus, le conflit apparaissant entre les sources d'informations fusionnées est mise en valeur et apportant ainsi une information pertinente sur la cohérence du système de fusion.

2.4.4. Bilan et choix du cadre de raisonnement pour la gestion de l'incertain

Le choix du cadre mathématique de l'incertain n'est pas toujours aisé. En effet, certains cadres sont plus ou moins privilégiés en fonction du domaine scientifique. Cette difficulté est bien présente dans le domaine multidisciplinaire de l'interprétation d'images. Quelques travaux ont tenté d'établir une comparaison entre les différents cadres en utilisant le cadre de la fusion d'information et de la reconnaissance d'objet [Bor00]. Les taux de reconnaissance obtenus définissent la performance de chaque cadre de raisonnement. Leurs conclusions montrent que les différents cadres réagissent de manière assez similaire.

Cependant, dans le cas d'informations proches du signal, issues des capteurs par exemple, les méthodes plus normatives, telle que l'approche probabiliste, sont généralement préférables. En revanche, les cadres de l'incertain plus récents tels que la théorie de l'évidence ou la théorie des possibilités sont plus flexibles dans les niveaux supérieurs de l'interprétation et robuste pour la gestion d'incertitude.

Dans le travail présenté ici, nous n'avons pas cherché à confronter ces approches, mais plutôt à les exploiter de manière opportuniste dans un premier temps. Cependant, dans un second temps, notre choix a été guidé par la généralité des fonctions de croyances et la flexibilité qu'offre le cadre de la théorie de l'évidence. Nous avons choisi la théorie des fonctions de croyance et en particulier le TBM qui est un cadre de travail générique basé sur les fonctions de croyance et proposant des outils performants pour la gestion de l'incertitude et de l'imprécision. Ce choix est aussi motivé par l'abstraction

des mécanismes de raisonnement du TBM qui distingue deux niveaux de raisonnement. Ce modèle de raisonnement abstraite permet de séparer la phase de représentation des connaissances appelée «*partie statique*» et celle de la fusion ou de la combinaison sur ces connaissances nommée «*partie dynamique*» de la phase de la prise de décision qui représente la «*partie décisionnelle*».

Conclusion

Nous avons présenté dans ce second chapitre un état de l'art des modules de traitements d'un système de vision pour la reconnaissance de scénario à travers sa structure modulaire. Les principales méthodes utilisées pour la reconnaissance d'événements d'objets mobiles dans les séquences vidéo ont été décrites. La diversité de ces techniques est un avantage mais aussi une difficulté car elle pose le problème de choix judicieux de l'approche à utiliser face à un type d'application. Les approches symboliques sont adaptées et flexibles pour des modèles de scénarios complexes et riches mais ne permettent pas de gérer l'incertitude. Les approches statistiques et probabilistes ont la capacité de gérer de manière très efficace les imperfections des données et celles issues des couches basses de l'interprétation.

Dans ce travail, nous avons exploité les approches probabilistes (modèle de Markov caché) et évidentielle (gestion de l'incertitude basée sur les modèles de croyances transférables MCT), en intégrant une couche de gestion de la structure hiérarchique des scénarios et une seconde couche de contraintes temporelles par une modélisation explicite de la durée du temps dans un état par le modèle. Le modèle proposé et exploité à trois niveaux d'abstractions: le premier est une hiérarchisation du modèle de Markov afin de gérer les sous-scénarios appartenant à la même catégorie. Puis le second niveau d'abstraction concerne la modélisation explicite de la durée de l'état du modèle afin de discriminer des scénarios de même niveau de l'hiérarchie mais qui se différencie par l'évolution temporelle des enchaînements des événements de base formant ces scénarios. Et enfin, le troisième niveau d'abstraction est relatif à la gestion de l'incertitude des données et des traitements de vision avec une prise en compte de l'incomplétude de connaissance (modélisation de l'ignorance de la connaissance experte par rapport à l'existence d'un scénario).

Chapitre 3

Reconnaissance de scénario par les Modèles de Markov Cachés Hiérarchiques et Temporels

Introduction

Dans ce chapitre, nous nous intéressons à l'analyse d'événements dans une scène dans le cadre de l'interprétation vidéo. A partir d'observations extraites sur la base de caractéristiques prédéfinies et obtenues en sortie du module de traitements bas niveau, une interprétation haut niveau de la scène est réalisée par exploitation d'outils mathématiques probabilistes. Nous avons choisi d'exploiter les modèles de Markov cachés afin de profiter de la puissance et de l'efficacité des algorithmes du modèle. Leur flexibilité, leurs diverses extensions, afin de répondre à une multitude d'applications sont des atouts majeurs qui ont guidé notre choix. Le modèle standard et des extensions adaptées ont permis, dans ce travail, de reconnaître des scénarios et des sous-scénarios avec une évolution temporelle d'état différente. Nous évoquons tout d'abord le modèle standard dans un contexte probabiliste, ensuite, exposons quelques extensions de ce modèle et nous nous attarderons sur le modèle hiérarchique. Enfin, nous présenterons une adaptation qui intègre un raisonnement temporel dans le modèle hiérarchique. Cette extension permet d'estimer de façon explicite le temps passé par le modèle dans un état par une variable probabiliste. Le modèle de Markov adapté proposé a été testé sur la séquence vidéo PEST'04 et a permis de reconnaître des scénarios complexes exploitables au sein d'une application de vidéosurveillance et spécifiquement en environnement médical qui constitue l'originalité de ce travail car la surveillance en USI n'avait jamais été jusqu'alors rendue visuelle et intelligente.

3.1 Les modèles de Markov cachés probabilistes

Les modèles de Markov cachés (*Hidden Markov Models : HMMs*) ont été introduits par Baum et al. à la fin des années 60 et bien étudiés au début des années 70 [Bau72], [Rab89]. Ce modèle est fortement apparenté aux automates probabilistes, définis par une structure composée d'états et de transitions, et par un ensemble de distribution de probabilité sur les transitions. A chaque transition des automates est associé un symbole d'un alphabet fini. Ce symbole est généré à chaque fois que la transition est empruntée. Un HMM se définit également par une structure composée d'états et de transitions et par un ensemble de distribution de probabilité sur les transitions. La différence essentielle avec les automates probabilistes est que la génération de symboles s'effectue sur les états et non sur les transitions. De plus, on associe à chaque état non pas un symbole, mais une distribution de probabilité sur les symboles de l'alphabet.

Les modèles de Markov cachés sont utilisés pour modéliser des séquences d'observations. Ces observations peuvent être de nature discrète (par exemple les caractères d'un alphabet fini) ou continue (fréquence d'un signal, température). Sans prétendre à une présentation exhaustive des modèles de Markov cachés, l'objectif de cette section est de dresser un portrait général et succinct de cette méthode et de son utilisation tant bien en apprentissage qu'en reconnaissance de séquences d'état. Les mécanismes et algorithmes présentés dans ce chapitre sont basés sur les travaux de L. Rabiner dans son tutoriel [Rab89]. Les trois algorithmes fondamentaux sont développés et les concepts de base sont présentés en Annexe B.

3.1.1. Présentation du modèle sur un cas d'école : balles dans des urnes

On dispose de 3 urnes contenant des boules blanches, noires et grises dans des proportions différentes. On note par $P(B/u_i)$, $P(N/u_i)$, $P(G/u_i)$ les probabilités de tirer une boule blanche, noire ou grise de l'urne u_i ($i=1,2,3$). Un opérateur invisible choisit, à des instants réguliers, une urne et tire une boule de cette urne et annonce la couleur. S'il choisit T urnes, il annoncera (à un observateur) la séquence des T couleurs tirés notée

O_T . Ainsi, au temps t , l'opérateur choisit une urne U_t et tire une boule de couleur O_t de cette urne. On suppose que le tirage d'une boule au temps t est indépendant du tirage de la boule au temps $t-1$. On fait l'hypothèse que l'opérateur choisit la séquence des urnes suivant un modèle d'une chaîne de Markov ayant dont les états sont les urnes. Cette chaîne sera dite cachée (puisque l'opérateur est invisible). En revanche, la séquence $U_T=U_1U_2\dots U_T$ sera dite la séquence des états cachée (car invisible) et elle est générée par la chaîne de Markov cachée. La séquence $O_T=O_1O_2\dots O_T$ sera dite la séquence des observations qui sera annoncée à l'observateur.

3.1.1.1. Définition et composantes du modèle

Les modèles de Markov sont des outils statistiques qui ont connu un grand succès ces trois dernières décennies du fait des multitudes applications dans lesquelles ils sont exploités. Ils sont connus comme l'un des outils statistiques les plus populaires qui ont été utilisés pour la modélisation et l'analyse de données séquentielles dans divers domaines d'applications.

Un modèle de Markov caché est un processus doublement stochastique avec un processus sous-jacent qui n'est pas observable. Il est représenté par un ensemble fini d'états, associé chacun à une distribution de probabilité (en général multidimensionnel). Les transitions entre les états sont régies par un ensemble de probabilités appelées probabilités de transition qui forment la matrice de probabilité de transition. Un état donné du modèle peut être généré par une observation, selon la distribution de probabilité associée. Mais lorsque l'état généré n'est pas visible pour un observateur externe, on dit qu'il est un état «caché»; d'où le nom modèle de Markov caché [Rab89]. Cinq principaux paramètres de base que sont S , O , A , B , Π permettent de définir complètement un modèle de Markov caché λ :

- S un ensemble de N états : $S = \{S_1, S_2, S_3, \dots, S_N\}$
- O un alphabet d'observations distinctes selon M symboles :

$$O = \{O_1, O_2, O_3, \dots, O_M\}$$

- A la probabilité de transition entre les états : $A = \{a_{i,j}\}$ où

$$a_{i,j} \equiv P(s_{t+1} = s_j | s_t = s_i)$$

- B la probabilité d'observations (probabilité d'émission, croyances) :

$$B = \{b_j(m)\} \text{ où } b_j(m) \equiv P(O_t = O_m | S_t = S_j)$$

- Π la probabilité d'état initial : $\Pi = \{\pi_i\}$ où $\pi_i \equiv P(S_1 = S_i)$

La Figure 3.1 illustre l'architecture d'un modèle de Markov caché à trois états.

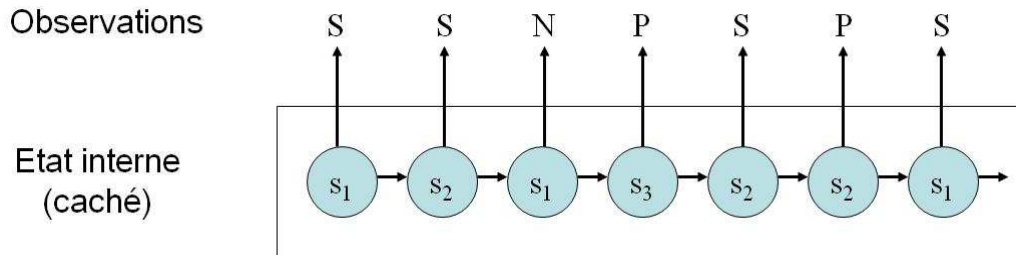


Figure 3.1. Architecture d'un modèle de Markov caché à trois états.

La séquence observée est l'évidence d'une chaîne de Markov cachée sous-jacente. L'émission de chaque état n'est pas déterministe. Chaque état peut émettre, de manière aléatoire, un parmi N symboles d'un alphabet. La Figure 3.2 montre le modèle des transitions et des composants du modèle.

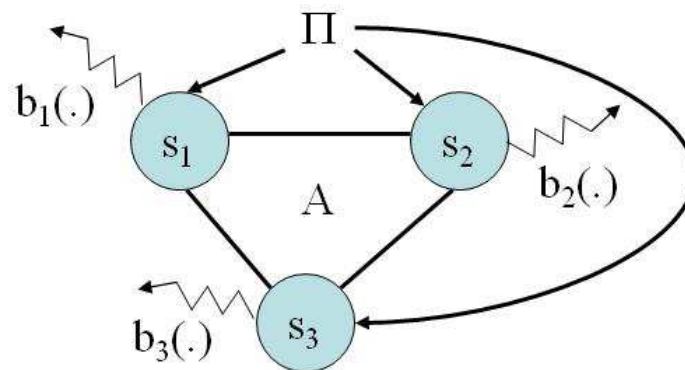


Figure 3.2. Architecture d'un modèle de Markov caché à trois états.

3.1.1.2. Problèmes de base et mécanismes de résolution

La procédure de génération d'une séquence $O = \{O_1, O_2, O_3, \dots, O_M\}$ de symboles à l'aide de modèle de Markov consiste, à partir d'un état S en suivant la distribution π , à se déplacer d'état en état en se basant sur les probabilités de transition, et à générer un symbole sur chaque état rencontré en utilisant la distribution de probabilité de

génération associée à l'état. Lorsqu'un symbole a été généré, on choisit une transition sortante suivant la distribution de probabilité de transition associée à l'état courant, et la procédure est réitérée jusqu'à la dernière génération de symbole de l'alphabet.

Les MMCs fournissent des solutions pour des applications d'apprentissage et de reconnaissance d'objets, d'une part et d'autre part de la classification et de segmentation de séquences, et en d'autres termes au découpage d'une séquence en sous-séquences de différents types. Pour ce faire, trois principaux problèmes de base caractérisant l'utilisation des modèles de Markov cachés sont généralement résolus par des algorithmes bien connus. Ces trois problèmes sont : l'évaluation du modèle, l'explication du modèle ou le décodage et enfin l'entraînement du modèle.

Le premier problème est un problème d'évaluation, qui peut également être vu comme un problème d'estimation de la capacité d'un modèle donné à reconnaître une séquence d'observation donnée. Le second problème se ramène à l'idée de dévoiler les états cachés S , sans y avoir accès directement. Dans la plupart des cas, le critère d'optimalité retenu influence la séquence d'états calculée. Enfin, le troisième problème se ramène à l'entraînement d'un modèle de Markov caché par des séquences d'observations, en vue d'en optimiser les paramètres pour un problème spécifique donné.

1. Le problème d'évaluation du modèle : étant donné un MMC λ , quel est le maximum de probabilité d'avoir une séquence d'observations $O = O_1, O_2, O_3, \dots, O_T$? L'estimation du maximum de probabilité se fait par calcul direct de la distribution probabiliste mais cette méthode fournit une solution très lourde en termes de complexité. Par conséquent, il faut utiliser l'algorithme *Forward-Backward* afin d'améliorer la complexité algorithmique.

2. Le problème d'explication ou le décodage: Etant donné un MMC λ et une séquence d'observations $O = O_1, O_2, O_3, \dots, O_T$, quelle est la séquence de la suite d'états dans λ qui a vraisemblablement généré O ? L'algorithme de *Viterbi* est une solution à ce problème.

3. Le problème d'apprentissage du modèle : Etant donné un jeu d'entraînement de séquences d'observations et un modèle initial λ du MMC, quel est le modèle λ du MMC qui aurait vraisemblablement généré le jeu. Cela revient à trouver comment ré-

estimer les paramètres de λ pour maximiser la vraisemblance de génération de cet ensemble de séquences? L'algorithme de **Baum-Welch** fournit la solution à ce problème.

Ces trois algorithmes classiques en solution aux problèmes de bases des MMCs sont détaillés dans les annexes dans leurs versions standards.

3.1.2. Typologie et topologie des modèles de Markov

A partir de la distribution des observations qui sont soit continues et infinies ou discrètes et finies après une transformation (quantification ou numérisation), nous distinguons principalement deux types de modèles de Markov cachés à savoir : le modèle continu qui est le modèle de base et le modèle discret obtenu après des traitements au préalable. Un autre modèle intermédiaire appelé modèle de Markov semi-continu a été développé en réponse aux imperfections des types de modèles. Nous y reviendrons dans la sous-section suivante consacrée à la présentation des extensions des modèles de Markov Cachés. Pour un modèle de Markov continu, la fonction de densité de probabilités (*Probability Density Function : PDF*) d'observations est continue. Ce qui est un avantage car elle est générée par des observations issues de phénomènes physiques qui, par conséquent, sont continues et permet de garder le signal (conservation de l'information contenu dans le signal sans dégradation) représentant le phénomène dans sa forme naturelle et plus proche de la réalité.

Cependant, cette forme des observations rend très ardues ou quasiment impossibles les traitements par une machine numérique. Chaque état du modèle continu intègre un mélange de fonctions de densité de probabilités représentant la probabilité d'observer un phénomène continu. La Figure 3.3 est un exemple de modèle de Markov à trois états de type continu où chaque état a un *PDF* de dimension un (1).

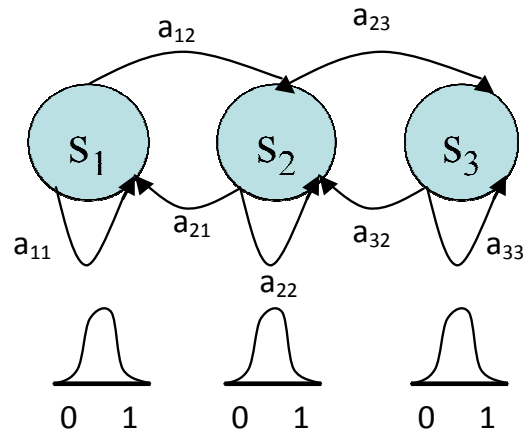


Figure 3.3. Exemple d'un modèle de Markov caché continu à trois états.

Dans le cas d'un modèle de Markov discret, les observations continues sont quantifiées par exemple à l'aide de symboles discrets d'un alphabet fini nommé O , et chaque état j a une fonction masse probabiliste qui définit la probabilité discrète que cet état émet un symbole O_m . Elle représente la probabilité d'observations, d'émissions ou de croyances nommé B et défini comme présenté à la section 3.1.1. Ainsi, contrairement au type continu, le modèle discret utilise une fonction de densité probabiliste discrète (*Probability Density Function: PDF*). Cette transformation peut entraîner des dégradations d'informations provoquées par la quantification des observations continues mais facilite et rend opérationnels des traitements informatiques temps-réel.

En vision par ordinateur, on utilise généralement un dictionnaire de mots à partir de symboles alphabétiques nommé *codebook*. Nous utiliserons ce terme *codebook* pour désigner ce dictionnaire par la suite tout au long de ce document. La Figure 3.4 est un exemple de modèle de Markov à trois états de type discret où chaque état a un *PDF* qui émet deux symboles 0 et 1.

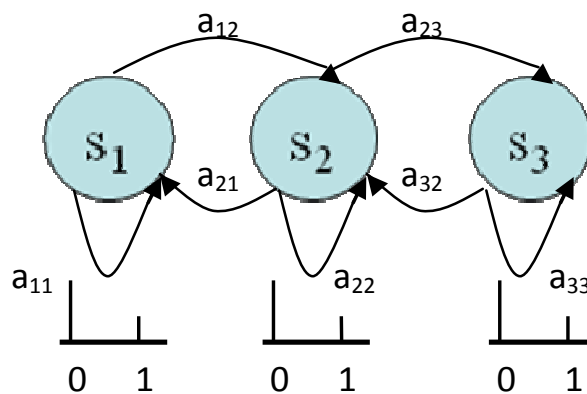


Figure 3.4. Exemple d'un modèle de Markov caché discret à trois états.

Différentes topologies des modèles de Markov cachés peuvent être envisagées en fonction de l'application pour laquelle il est utilisé pour la modélisation. Nous ferons état de trois topologies bien connues dans notre travail.

1) La topologie ergodique où chaque état peut être atteint à partir de n'importe quel autre état du modèle en un nombre fini de transitions et est caractérisée par une matrice de transition T aux coefficients strictement positifs (Figure 3.5).

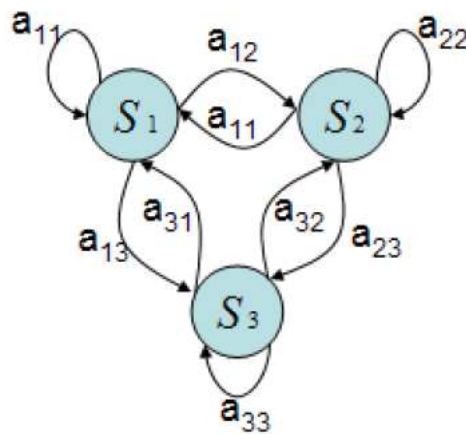


Figure 3.5. Exemple d'un modèle de Markov caché ergodique à trois états.

2) La topologie issue des *modèles de Bakis* appelée encore *topologie gauche-droite* n'autorise aucune transition d'un état à un autre. Les états se succèdent et ont des indices égaux ou supérieurs aux précédents. Dans ce cas, la probabilité initiale du premier état est fixée égale à un, les autres étant mises à zéro. Certaines valeurs des probabilités de transitions sont mises à zéro et les transitions entre états sont acceptées si et seulement si $a_{ij} > 0$. En conséquence, le modèle démarre toujours et se termine toujours par le dernier état pour un élément appartenant au système. Cette topologie est divisée en deux catégories à savoir la topologie séquentielle et celle parallèle.

Nous présentons ici la topologie séquentielle qui contient moins de paramètres et est bien adaptée pour la modélisation d'enchaînements d'événements dont l'évolution suit un ordre bien connu suivant la procédure suivante : à partir de l'état initial, on se déplace d'état en état suivant les probabilités de transition jusqu'à atteindre l'état final et pour chacun des états rencontrés, on émet un symbole en fonction de la distribution des probabilités associées. Cette structure du modèle

permet d'émettre des séquences de symboles sans une estimation au préalable de densité de probabilité initiale, ce qui est un avantage de la topologie. La Figure 3.6 est un exemple de topologie gauche-droite séquentielle du modèle de Markov à trois états.

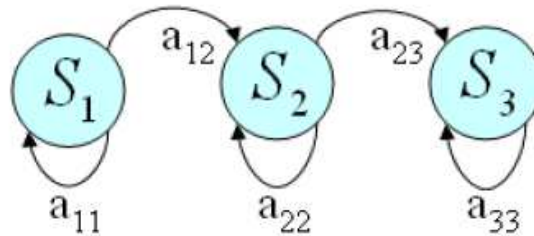


Figure 3.6. Exemple de topologie gauche-droite séquentielle d'un MMC à trois états.

3) Le modèle gauche-droite parallèle où chaque état peut être atteint à partir de n'importe quel autre état du modèle en un nombre fini de transitions permet de prendre en compte la possibilité de multiples successions ou enchaînements de symboles pour le même événement. Par exemple, en reconnaissance de parole, cette topologie a été très utilisée pour modéliser le même mot avec plusieurs prononciations. D'autres topologies des modèles de Markov existent mais nous nous limiterons à la présentation de ces trois topologies de base. La Figure 3.7 illustre la topologie d'un MMC gauche-droite parallèle à six états.

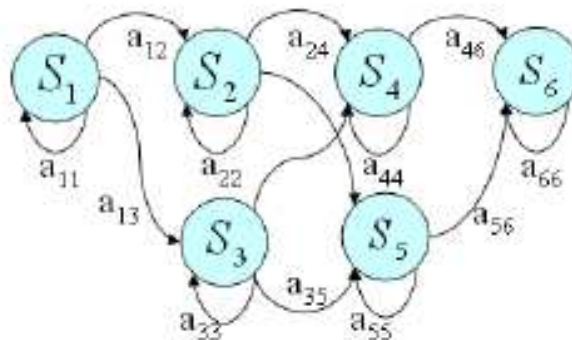


Figure 3.7. Exemple la topologie gauche-droite d'un MMC parallèle à six états.

3.2. Les extensions du modèle standard

De nombreuses extensions des MMCs ont été proposées pour répondre aux différentes limitations du modèle standard bien que simple et efficace. Ainsi, pour

modéliser au mieux la nature réelle des données par un MMC, il est opportun d'intégrer des contraintes spécifiques dans la définition des modèles. Ces contraintes spécifiques sont définies par rapport à une connaissance a priori sur le phénomène à modéliser [PH01].

Une première contrainte intégrée dans le modèle MMC de base est une structure hiérarchique [FST98] permettant de représenter des séquences d'observations à divers niveaux d'abstraction et cela répond à des exigences de plusieurs applications dans des domaines très variés tels que la reconnaissance de parole, la reconnaissance de séquence biologique, la reconnaissance des scènes et autres. Les études sur la reconnaissance vocale ont permis l'édification de nombreux modèles spécialisés. La structure fortement hiérarchisée de la parole et les riches études sur le domaine ont rendu cette édification possible et performante.

Une seconde contrainte concerne la durée d'état. Elle consiste à estimer de façon explicite la durée passée par le modèle dans un état donné qui s'avère un paramètre très important dans le cadre de la modélisation de données séquentielles fortement temporelles. Ce second type de contrainte a donné naissance aux Modèles Semi Markovien Cachés (MSMC) qui ont été un sujet de recherche très actif depuis la fin des années 1980. En effet, plusieurs applications dans le domaine du traitement et la reconnaissance vocale, de la reconnaissance de caractères en écriture et de la reconnaissance de scénario dans une vidéo constituent le succès et l'émergence des extensions. Cette forme d'extension répond au problème de l'hypothèse de Markov (distribution exponentielle ou géométrique de la durée des états dans le MMC) dans la modélisation des observations qui permet de gérer la distribution non exponentielle de la durée des états (ou non géométrique si le temps est discret).

Par exemple, en 1985, Russel et Moore ont utilisé une distribution de probabilité de type Poissonniène pour modéliser la durée de l'état de façon explicite dans le contexte d'une application de reconnaissance de la parole [RJM95]. De même, en 1986, Levinson [Lev86] a proposé, dans le cadre de la reconnaissance de la parole, un MMC intégrant une modélisation de la durée du temps passé par le modèle Markovien dans un état à base de la distribution continue de probabilité de type Gamma et de type gaussien.

Une des premières extensions développées fut le MMC factoriel, dit en anglais *Factorial Hidden Markov Model (FHMM)*, qui a permis de représenter différents niveaux du langage (phonèmes, syllabes, mots, phrases). Pourtant, bien que Logan et Moreno [LM98] aient indiqué que cette modélisation n'augmentait pas les performances générales en matière de reconnaissance vocale, elle montre son efficacité quand le signal à traiter est issu de plusieurs processus indépendants [LM98], [RGR03]. Dans le cadre de cette thèse, nous nous sommes concentrés sur une étude des travaux connexes sur la reconnaissance des activités humaines dans une vidéo exploitant des extensions des MMCs.

Dans la littérature, la plupart des travaux de recherche sur la thématique de reconnaissance d'activités humaines dans des séquences vidéo se sont concentrés sur la représentation et l'apprentissage des caractéristiques séquentielles et temporelles décrivant l'activité à reconnaître. A cette fin, l'utilisation de modèles dynamiques très répandus tels que le MMC [HJB98, ABT10] a fait le premier succès de ces applications en vision par ordinateur. Le MMC standard est un modèle simple et efficace pour l'apprentissage et la reconnaissance des données séquentielles. En outre, son rendement tend à se dégrader lorsque l'éventail des activités devient de plus en plus complexe, ou lorsque les activités à reconnaître intègrent de forte dépendance temporelle à long terme qui reste suffisamment difficile à modéliser et à traiter avec l'hypothèse de base de Markov. Face à ces deux caractéristiques spécifiques que sont l'ordre d'évolution séquentielle des actions élémentaires formant l'activité qui peut être étendue à la complexité de la structure architecturale (par exemple l'existence d'une hiérarchie) de l'activité et l'évolution temporelle des actions de base permettant de décrire l'activité, deux grandes catégories d'extensions sont proposées par les chercheurs.

La première catégorie d'extension qui complète le MMC de base introduit une structure hiérarchique dans le modèle standard, visant à exploiter l'organisation hiérarchique naturelle des comportements humains. Dans cette optique d'idée, le modèle de MMC appelé *Abstract HMM* [BVW02] a été proposé de même qu'un type hiérarchique (*Hierarchical Hidden Markov Model: HHMM*) [PH01, FST98], et le modèle de Markov caché à couches (*Layered Hidden Markov Modèles*) [OHG02]. La dépendance

séquentielle à long terme est gérée dans ces modèles via les couches supplémentaires visant à modéliser l'évolution des échelles de temps plus lents des activités à un haut niveau d'interprétation.

La seconde catégorie d'extension du modèle de MMC classique adopte le modèle semi-Markov et présente deux grandes variantes que sont les modèles de Markov à durée d'état explicite (*Hidden Semi-Markov Model : HSMM*) [DH07, Mur02] qui intègrent un paramètre ou une variable qui permet de quantifier de façon explicite la durée passée par le modèle dans un état et le modèle de Markov segmentaires (*Segmental Hidden Markov Model : HMM segmental*) [BDH07]. Dans ces modèles, un état est supposé inchangé pour une certaine durée de temps aléatoire avant son passage à un nouvel état. Pour chaque état, une distribution de la durée est donnée pour caractériser la longueur de cette dernière. Comme nous l'avons vu précédemment plus haut en début de cette section, les MMC classiques sont très efficaces en classifications de séquences temporelles, et en particulier lorsque les scénarios sont basés sur une succession d'événements avec une granularité temporelle constante. Pour des scénarios plus complexes, dont la complexité est liée à la structure organisationnelle et l'évolution temporelle des séquences, il est nécessaire d'intégrer diverses extensions permettant d'enrichir le modèle de base.

Dans la section suivante, nous nous intéressons aux modèles hiérarchiques [FST98, RJM85] et modèles semi-Markovien [BVW01, LVW⁺04a, LVW⁺04b] pour la reconnaissance d'événement d'objets puis déboucher sur une proposition d'approche hybride de ces deux extensions afin de gérer d'une part la structure hiérarchique des scénarios et d'autre part, l'estimation explicite de la durée d'état (un événement élémentaire) via une variable probabiliste. Enfin, une exploitation de ce modèle pour la reconnaissance d'événements dans des séquences vidéo est proposée au chapitre 5 avec une analyse du taux de reconnaissance obtenu.

3.3. Les Modèles de Markov Cachés Hiérarchiques et Temporels

Après avoir décrit les MMCHs et le modèle semi-markovien caché, nous présentons un modèle hybride basé sur la combinaison des deux extensions de base présentées que nous appelons Modèle de Markov Caché Hiérarchique et Temporel (MMCHT). Les différents mécanismes mis en œuvre par cette extension y sont aussi présentés. L'algorithme de *Baum-Welch* modifié intégrant une contrainte temporelle permet d'assurer l'étape d'apprentissage des séquences temporelles et une gestion de leur structure hiérarchique. De même, pour l'étape de reconnaissance, les mécanismes de propagation avant et arrière implémentés par l'algorithme de *Forward-Backward* modifié de type temporel ont été aussi présentés.

3.3.1. Modèle de Markov Caché Hiérarchique (MMCH)

Les MMCHs ont été introduits comme une extension des MMCs standards par Fine, Singer et Thisby dans les années 1998 [FST98]. Cette extension a été inspirée de la propriété des langages réguliers qui sont modélisés à base d'automates à états finis (AEF) qui sont ensuite transformés en un arbre hiérarchique. Plus spécifiquement, un MMCH est une hiérarchie où le nombre de niveaux hiérarchiques augmente de plus haut niveau vers le niveau le plus bas, et où les observations générées dans un état I_k appartenant à un automate stochastique au niveau k sont des séquences générées par un automate au niveau $K+1$.

Les MMCHs constituent une extension des MMCs où une séquence peut contenir une sous-séquence elle-même modélisée par un MMC ou un MMCH. Un état parent émet une suite de symboles plutôt qu'un seul symbole, et cela par activation de ces états fils (sous-états). Ces états fils peuvent être à leur tour composés de MMC ou de MMCH et ainsi de façon récursive. Les états n'émettant pas directement des symboles sont des états appelés «*états internes*». Ainsi la structure d'un MMCH permet de corréler directement des événements qui se produisent à divers niveaux dans une séquence d'observation. De façon générale, deux types de transitions sont définis en fonction du niveau auquel la liaison est réalisée: l'un appelé transition «*verticale*» et l'autre

«horizontale». Une transition «verticale» permet de relier un état interne à ces états fils tandis qu'une transition «horizontale» permet de relier deux états de même niveau.

3.3.1.1. Définition et topologie d'un MMCH

Les MMC Hiérarchiques sont des MMC qui permettent de par leur flexibilité de représenter hiérarchiquement une séquence d'état et par conséquent offrant par ce biais une solution très intéressante pour la représentation des séquences d'observation à différents niveaux d'abstraction. En effet, au cours des mécanismes génératifs, chaque état interne peut activer un sous-modèle correspondant. L'activation d'un sous-modèle met en évidence une transition verticale. La fin de cette procédure récursive est réalisée lorsque des états terminaux spéciaux, appelés états de production, sont atteints. Ces états n'ont pas de sous-modèle correspondant et sont les seuls qui génèrent un symbole par le mécanisme de sortie d'un MMC classique. A la fin d'une transition verticale, le système revient à l'état d'origine qui a généré la séquence de transitions récursives. La Figure 3.8 illustre la structure d'un MMC à topologie hiérarchique.

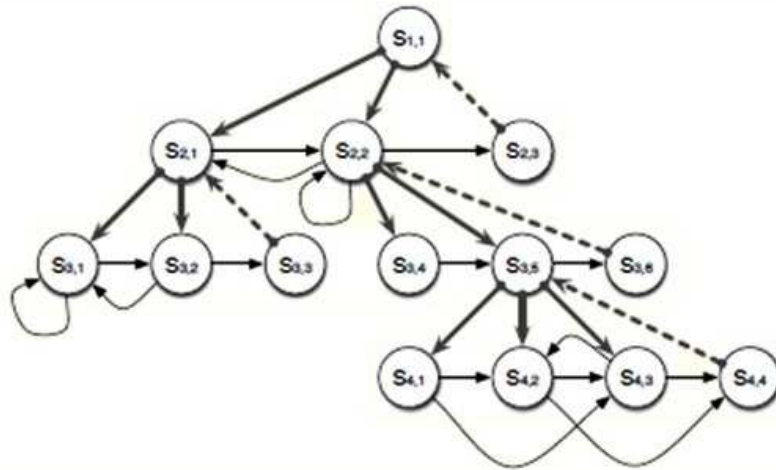


Figure 3.8. Exemple d'un modèle de Markov caché Hiérarchique.

3.3.1.2. Modélisation et composantes du modèle MMCH

Les composantes d'un MMCH sont définies comme dans le cas d'un MMC classique. Pour chaque sous-automate (représenté par un MMC standard), les transitions d'un état à un autre sont régies par une distribution A , et la distribution de probabilité pour un état initial est régie par une loi de distribution notée Π . Cependant,

il existe une restriction sur le nombre d'états terminaux qui est fixé à 1. Les seules transitions horizontales autorisées sont celles entre les états du même sous-modèle et en d'autres termes du même sous-automate, autrement dit, aucune transition directe ne peut se produire entre les différents états des automates dans la hiérarchie. En considérant un MMCH noté λ^n où $n \in [1, \dots, N]$, n représente le niveau de la hiérarchie.

Nous désignons un état générique S_i^n (état émetteur ou état interne) avec i l'indice de l'état (S_i^n est le $i^{\text{ème}}$ état du niveau n de la hiérarchie). λ^n peut être complètement défini

et formalisé par quatre principaux paramètres de base que sont : $S = \{S_i^n\}$, $A^{S^n} = \{a_{i,j}^{S^n}\}$, $B = \{B^{S^n}\}$, $\Pi = \{\Pi^{S^n}\}$.

- S un ensemble de N états : $S = \{S_i^n\} = \{S_1^n, S_2^n, S_3^n, \dots, S_N^n\}$, n désignant le niveau de la hiérarchie, elle varie de 1 (la racine) à N pour les états de production. Le modèle d'observation est conçu au niveau $n=N$ de la hiérarchie et pour chaque état générique S_i^n on note $|S_i^n|$ le nombre de sous-états de cet état générique.

- A représente la probabilité de transition entre les états internes : $A^{S^n} = \{a_{i,j}^{S^n}\}$ où $n \in [1, \dots, N-1]$ et $a_{i,j}^{S^n} = P(S_j^{n+1} | S_i^{n+1})$

- B la probabilité d'observations (probabilités d'émission, croyances) : $B = \{B^{S^n}\} = \{b^{S^n}(k)\}$ où $k \in [1, \dots, M]$ tel que $b^{S^n} = P(O_k | S^n)$ et représente la probabilité d'observer le symboles O_k .

- O représente l'ensemble discret des M symboles du vecteur d'observation associé aux états émetteurs tel que : $O = \{O_1, O_2, O_3, \dots, O_M\}$

- Π la probabilité d'état initial : $\Pi = \{\Pi^{S^n}\} = \{\pi^n(S_i^{n+1}) = P(S_i^{n+1} | S_i^n)\}$.

π^{S^n} est aussi appelé probabilité de transition verticale entre S^n et S^{n+1} . Il faut noter que la probabilité qu'un état terminal (état de sortie) noté S_f^{n+1} termine une transition verticale noté $\Gamma^{S^n} O_k^n = P(S^n | S_f^{n+1})$

La Figure 3.9 montre un MMCH avec une topologie hiérarchique à trois niveaux. Les états grisés sont des états internes encore appelés états parents. Les flèches en pointillées mènent vers l'état de sortie (état final). Par exemple, l'état interne 2 du niveau 2 possède $N=2$ sous-états et caractérisée par les matrices suivantes :

$$S^{s_2^2} = \{s_1^3, s_2^3\} \text{ et } A^{s_2^2} = \begin{pmatrix} 0.7 & 0.3 \\ 0 & 0 \end{pmatrix} \text{ puis } \pi^{s_2^2} = (1 \ 0) \text{ et enfin } \Gamma^{s_2^2} = (0 \ 1)$$

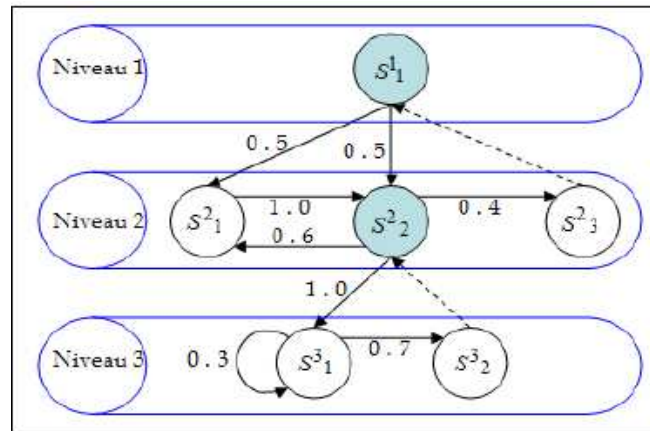


Figure 3.9. Exemple d'un MMCH à trois niveaux [Zia10].

3.3.2. Modèle Semi-Markovien Caché avec durée d'état explicite

L'utilisation des MMC classiques de même que celle du modèle hiérarchique précédemment étudié ne permet pas de tenir compte de la durée variable des événements de base et par conséquent de leur enchaînement (activité ou scénario à reconnaître). Cette lacune du modèle Markovien standard peut constituer un frein pour un bon taux de reconnaissance et mieux le paramètre temporel peut faire office d'un bon attribut qui permet de discriminer les différents événements d'une scène vidéo. C'est le cas de notre application de surveillance intelligente en milieu médical. Pour une description plus fine des événements (scénarios), il convient de trouver une solution pour modéliser explicitement et estimer la quantité de temps qu'utilise le modèle MMC tout en demeurant dans un état donné.

Dans la littérature, une extension du MMC standard à celui semi-Markovien fournit des solutions à ce problème. Dans la section suivante, nous exposerons le modèle semi-Markovien puis le modèle proposé qui est un MMC Hiérarchique et Temporel (MMCHT).

Dans le modèle de MMC standard, l'estimation de la durée de l'état est implicite. Elle est en fonction de la probabilité d'auto-transition d'un état du modèle. Toujours, dans le cadre du modèle classique, la probabilité d'occupation d'un état du système est a priori supposée distribuée géométriquement. Cette distribution de la durée d'état n'est pas adéquate si les états modélisent de façon significative des traits temporels. Soit i un

état du système et a_{ii} la probabilité auto-transition (une boucle sur l'état) de cet état, nous pouvons montrer que la probabilité de rester dans cet état i pour d «pas» de temps consécutifs est exponentielle et estimée par l'équation 3.1.

$$(a_{ii})^{d-1} \times (1 - a_{ii}) \quad (3.1)$$

Le Modèle Semi-Markovien Caché (MSMC) [RJM95]) a été développé pour permettre le paramétrage explicite de la durée d'état et l'inférence directe des distributions de durée de l'état. La quantification de la durée d'état explicite [DVB+05, DVB+09, LVW+04a, LVW+04b] du modèle temporel passe par l'estimation d'un paramètre de type probabiliste noté $P_i(d)$ tel que $1 \leq d \leq D$ où D définit la limite de la durée maximale que peut passer le modèle dans un état. P_i est un vecteur de probabilités de durée discrète tels que $\sum_{d=1}^D p_i(d) = 1$ et les probabilités des transitions indépendantes sont liées de sorte que $a_{ii} = 0$.

3.3.3. Modélisation de la durée d'état dans un MMCHT

Un MMC avec contrainte temporelle noté $\lambda = (A, B, \Pi, N, P_j(d))$ où $P_j(d)$ représente la probabilité que le MMC passe un temps de durée d dans l'état j est estimée avec des distributions probabilistes qui correspondent à l'évolution temporelle de la séquence d'états du modèle et estimée statiquement. N états formant un ensemble S , O un alphabet d'observations distinctes selon M symboles, A la probabilité de transition entre les états, B la probabilité d'observations, la probabilité d'état initial. Deux principales classes de distributions peuvent être utilisées: la distribution discrète et la distribution continue.

3.3.3.1. Distribution Discrète pour la durée d'état d'un MMCH

Une distribution de Poisson a été utilisée par Russel [RJM85] pour modéliser la durée dans un MMC. Dans ce cadre, la probabilité d'une durée d dans l'état j est donnée par l'équation 3.2:

$$P_j(d) = e^{-l_j} \times \frac{(l_j)^d}{d!} \quad (3.2)$$

La variable aléatoire d désigne la durée passée dans l'état j , qui suit cette distribution, et a pour espérance l_j , qui est le paramètre de la loi de Poisson. L'estimation de l_j s'effectue par calcul de l'espérance de la durée passée dans l'état divisée par l'espérance de l'occurrence de cet état. Par exemple, l'équation 3.3 permet d'estimer le paramètre l_j où $X_{t_0,t_1}(j)$ est obtenu avec l'équation 3.4.

$$l_j = \frac{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j) \times (t_1 - t_0 + 1)}{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j)} \quad (3.3)$$

$$x_{t_0,t_1}(j) = \frac{\sum_{i=1, i \neq j}^N \alpha_{t_0,t_1}(j) \times a_{ij} \prod_{s=t_0}^{t_1} b_j(o_s) \times p_j(t_1 - t_0 + 1) \times \beta_{t_1}(j)}{P(O/\lambda)} \quad (3.4)$$

3.3.3.2. Distribution continue pour la durée d'état d'un MMCHT

Dans le cadre de la reconnaissance automatique de la parole, Levinson et al. ont proposé deux distributions continues que sont la distribution gamma et la gaussienne, pour modéliser explicitement la durée dans les modèles de Markov cachés [Lev86]. Du fait que ce sont des lois continues, ces modèles ont été dénommés MMC à durée variable continue et le terme consacré en anglais "*Continuous Variable Duration HMMs*" est le plus connu et utilisé [Lev86].

- **Distribution gaussienne**

La densité de durée d'état modélisée par une distribution gaussienne s'écrit :

$$P_j(d) = \frac{1}{\sigma_j(2\pi)^{1/2}} \times e^{-\frac{(d-m_j)^2}{2\sigma_j^2}} \quad (3.5)$$

où m_j et σ_j désignent respectivement la moyenne et la variance de la loi gaussienne.

• **Distribution Gamma**

Lorsque la densité de durée d'état est modélisée par une distribution Gamma, elle s'écrit [BEM06]:

$$P_j(d) = \frac{\eta_j^{v_j} d^{v_j-1} e^{-\eta_j d}}{\Gamma(v_j)} \tag{3.6}$$

avec les paramètres η_j, v_j et $\mu_j = v_j^{-1}$ la moyenne et $\sigma_j^2 = v_j \eta_j^{-2}$ la variance. L'équation permet d'estimer la variance où le paramètre ξ_j est obtenu en utilisant l'équation 3.7.

$$\sigma_j = \frac{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j) \times (t_1 - t_0 + 1)^2}{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j)} - (\xi_j)^2 \tag{3.7}$$

$$\xi_j = \frac{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j) \times (t_1 - t_0 + 1)}{\sum_{t_0=1}^T \sum_{t_1=t_0}^T x_{t_0,t_1}(j)} \tag{3.8}$$

3.3.4. Mécanismes de ré-estimation des paramètres d'un MMCH

Les formules du modèle de ré-estimation sont présentées dans les équations 3.10 et 3.12, compte tenu des variables de *forward* α et *backward* β dans les équations 3.9 et 3.11 et la séquence d'observations $O_1, O_2, O_3, \dots, O_{t-1}, O_t$ produit au fil du temps t tel avec $1 < t < T$ et $O_t \in V$ où $V = \{V_1, V_2, \dots, M_V\}$ représente l'ensemble des symboles d'observations possibles [SSG+04a, SSG+04b].

$$\alpha_t(j) = P(O_1, O_2, O_3, \dots, O_t, S_j \text{ fin à } t/\lambda) \tag{3.9}$$

$$\alpha_t^*(j) = P(O_1, O_2, O_3, \dots, O_t, S_j \text{ début à } t+1/\lambda) \tag{3.10}$$

$$\beta_t(j) = P(O_{t+1}, \dots, O_T / S_i \text{ fin à } t, \lambda) \tag{3.11}$$

$$\beta_t^*(j) = P(O_{t+1}, \dots, O_T / S_i \text{ début à } t+1, \lambda) \tag{3.12}$$

$$\pi_i^* = \frac{\pi_i \beta_0^*(i)}{P(O|\lambda)} \quad (3.13)$$

$$a_{i,j}^* = \frac{\sum_{t=1}^N \alpha_t(i) a_{i,j} \beta_t^*(j)}{\sum_{j=1}^N \sum_{t=1}^T \alpha_t(i) a_{i,j} \beta_t^*(j)} \quad (3.14)$$

3.3.5. Mécanismes d'apprentissage et de reconnaissance de scénarios par exploitation d'un MMCHT

Nous décrivons dans cette section les composantes et les différents mécanismes du Modèle de Markov Caché Hiérarchique et Temporel (MMCHT) que nous proposons et exploitons pour reconnaître un scénario dans des vidéos médicales. L'approche proposée est basée sur MSMC et intègre une gestion explicite de la durée d'état du modèle par l'estimation d'une variable probabiliste via des distributions discrètes ou continues. En effet, le modèle est composé d'une suite de tranche de temps, où chaque laps de temps est en fait une variable d'activation de la fin de l'état courant, et la même structure de réseau est répliquée au fur et à mesure que le temps avance. Les variables temporelles sont reliées entre elles par des arcs de la manière dont les états courants sont déclenchés par la fin de la variable précédente. Ce modèle permet de reconnaître des activités se retrouvant à un même niveau de la structure et par la même occasion, nous offre la possibilité d'estimer les temps passés par l'automate dans un état donné afin de discriminer temporellement ces scénarios se retrouvant au même niveau de la hiérarchie.

Un MMCHT est formalisé par l'ensemble $\lambda = (A^S, B^S, \pi^S, P_s(d), N)$ où $P_s(d)$ est la probabilité que le modèle soit resté dans un état émetteur pendant une durée d . les algorithmes de *Baum-Welch* et *Backward-Forward* de base ont été modifiés pour prendre en compte la structure hiérarchique et la distribution du temps de séjour dans un état.

L'apprentissage des paramètres du modèle à travers l'estimation et l'ajustement des paramètres du modèle précède l'étape de reconnaissance proprement dite. A cet effet, nous avons utilisé l'algorithme de *Baum-Welch* à durée d'état explicite d'un MMCH qui se présente comme suit :

Algorithme 1 : Procédure de Baum-Welch hiérarchique et temporelle

- 1: Initialisation choix des paramètres du MMC
 - 2: Répéter :
 - 3: **Pour** chaque séquence observée et à un niveau n de la hiérarchie:
 - 4: **Calculer** $\alpha(s_j^n)$ avec une procédure forward hiérarchique temporelle
 - 5: **Calculer** $\beta(s_j^n)$ avec une procédure backward hiérarchique et temporelle
 - 6: **Calculer** $\xi(s_j^n)$
 - 7: **Calculer** $\gamma(s_j^n)$
 - Fin pour**
 - 8: Ré-estimer les paramètres du modèle jusqu'à ce qu'ils deviennent stables.
-

Une fois l'étape d'apprentissage des paramètres effectuée du modèle, la reconnaissance des séquences peut s'effectuer en exploitant l'algorithme de retour Avant-Arrière (*backward-forward*) à durée d'état explicite d'un modèle de Markov caché hiérarchique qui se présente comme suit :

Algorithme 2 : Procédure forward hiérarchique et temporelle $\alpha(s_j)$

- 1: Initialisation du niveau n de la hiérarchie
 - 2: **Pour** i allant de 1 à N faire $\alpha_1^n(i) = \pi_i b_i(O_1) P_1(1)$
 - 3: **Fin pour**
 - 4: **Pour** t allant de 1 à $T-1$ faire
 - 5: **Pour** i allant de 1 à N faire
 - 6: **Fin pour** $\alpha_{t+1}^n(i) = \left[\sum_j \alpha_t(j) a_{ji} \right] b_i(O_{t+1})$
 - 7: **Fin pour**
 - 8: **Calculer :** $P(O|\lambda) = \sum_{i=1}^N \alpha_t(i)$
-

Le modèle d'estimation de la probabilité $P(O|\lambda)$ de vraisemblance d'un MMC λ à durée d'état explicite, pour une observation séquence de variable S , peut être calculé par un algorithme *forward-backward* à partir de l'équation 3.15:

$$P(O / \lambda) = \sum_{i=1}^N \sum_{j=1, i \neq j}^N \sum_{d=1}^t \alpha_{t-d}(i) a_{ij} P_j(d) \prod_{s=t-d+1}^t b_j(o_s) \beta_t(j) \quad (3.15)$$

Algorithme 3 : Procédure backward hiérarchique et temporelle $\beta(s_j)$

1: Initialisation du niveau n de la hiérarchie

2: **Pour** i allant de 1 à N faire : $\beta_t(i) = 1$

3: **Fin pour**

4: **Pour** t allant de 1 à $T-1$ faire :

5: **Pour** i allant de 1 à N faire:

$$\beta_t(i) = \sum a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

6: **Fin pour**

7: **Fin pour**

8: **Calculer :** $P(O / \lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i)$

3.4. Application à la reconnaissance des scénarios «utiliser un distributeur» et «bagage abandonné»

Dan cette section, nous nous proposons de tester notre algorithme pour la reconnaissance des scénarios «utiliser un distributeur» et «bagage abandonné» dans le cadre de la surveillance des événements dans un hall d'entrée d'une gare. La séquence que nous utilisons est celle bien connue exploitée par la communauté de vision par ordinateur nommée PETS'2004. Nous définissons la structure du hall d'entrée de la gare comme constituée de zones d'intérêts. Chaque zone est elle-même décomposée en événements auxquels elle est associée. Les connaissances à priori que nous disposons sur l'environnement nous permettent de modéliser la structure et définir le nombre d'états du modèle MMCHT qui permettra de reconnaître les scénarios.

Nous formalisons le modèle du MMCHT par $\lambda = (A^S, B^S, \pi^S, \Gamma^S, P_S(d), N)$ où $P_S(d)$ est la loi probabiliste du temps de séjour dans l'état émetteur S et d la durée de l'état du modèle. La MMCHT utilisé pour modéliser l'automate probabiliste qui permettra de reconnaître le scénario «utiliser un distributeur» dans un hall d'entré de la gare a une structure arborescente à une seule racine au sommet de la hiérarchie comme le montre

la Figure 3.10. Les états émetteurs sont des feuilles et les liens entre les états sont des branches. Ainsi pour reconnaître les événements dans une scène du hall d'entrée d'une gare, nous modélisons le fait qu'un piéton rentre dans le hall se dirige vers la zone d'un guichet pour l'utiliser (achat d'un titre de voyage par exemple) avant de sortir du hall par un modèle de Markov caché hiérarchico-temporel λ illustré à la Figure 3.10.

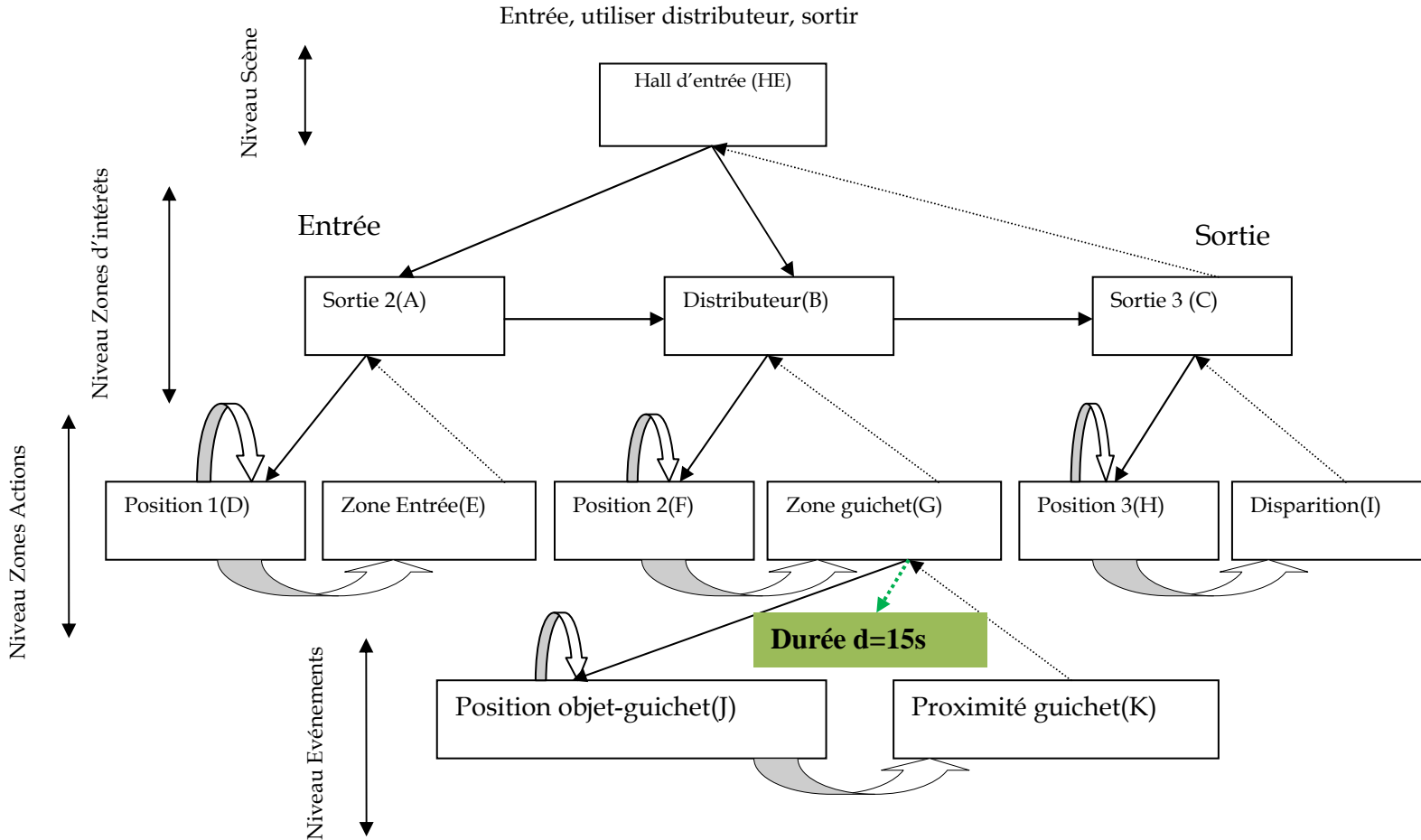


Figure 3.10. Structure hiérarchique du MMCHT exploité pour reconnaître le scénario «utiliser un distributeur».

Dans la scène étudiée, le MMCHT utilisé pour reconnaître les événements est composé de quatre niveaux hiérarchiques : scène, les zones d'intérêts, niveau zones-actions et enfin niveau actions. On choisit n le niveau de la hiérarchie et ainsi chaque niveau est décrit comme suit :

Au niveau «scène» où $n=1$, le MMCHT est constitué d'un seul état e noté S_1^1 . Cet état représente la racine de la hiérarchie et est composé de trois sous-états appartenant au niveau deux qui est celui des zones d'intérêts que sont sortie 2 (noté A), distributeur

(noté B), sortie 3 (noté C). Les matrices de transition horizontales et verticales de l'état S_1^1 sont fixées a priori à :

$$A^{S_1^1} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad \Pi^{S_1^1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{et} \quad \Gamma^{S_1^1} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Au niveau «zones d'intérêts» où $n=2$, le MMCHT est constitué de trois états noté $\{S_1^2, S_2^2, S_3^2\}$ qui sont des zones d'intérêts prédéfinis. L'état S_1^2 est composé de deux sous-états labellisés respectivement (D) et (E) de niveau «zones d'intérêts» que sont la zone d'entrée et la position en position de l'objet par rapport à la sortie 2. De même l'état S_2^2 est composé de deux sous-états labellisés respectivement (F) et (G) de niveau «zones d'intérêts» que sont la zone guichet et la position en position de l'objet par rapport au distributeur et enfin l'état S_3^2 est composé de deux sous-états labellisés respectivement (H) et (I) de niveau «zones d'intérêts» que sont la zone de disparition de l'objet et la position de l'objet par rapport à la sortie 3. Les matrices horizontales et verticales associées sont :

$$A^{S_1^2} = A^{S_3^2} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}, \quad \Pi^{S_1^2} = \Pi^{S_3^2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{et} \quad \Gamma^{S_1^2} = \Gamma^{S_3^2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$A^{S_2^2} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Pi^{S_2^2} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{et} \quad \Gamma^{S_2^2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Au niveau «zone-actions» où $n=3$, le MMCHT est constitué de six états noté $\{S_1^3, S_2^3, S_3^3, S_4^3, S_5^3, S_6^3\}$ de niveau «zone-actions» modélisant l'entrée et la sortie de l'objet observée dans la scène du hall d'entrée de la gare. Dans un tel hall plusieurs événements peuvent être détectés et peuvent être qualifiés d'anormaux ou pas mais nous nous contenterons de tester notre technique pour reconnaître un scénario dans lequel une personne entre dans le hall peut faire un tour et va utiliser le distributeur avant de sortir du hall. Il faut noter que la reconnaissance du scénario étudié est sur la détection de personne à proximité du distributeur (action de base proximité du guichet) puis de sa présence effective au niveau du distributeur (position de l'objet par rapport au guichet) et enfin le temps passer par ce dernier au niveau du distributeur représenté

par la probabilité $P_G(d)$. La durée en unité de mesure de temps est supposée $d=30s$, $d=60s$ ou $d=120s$) et l_s représente le paramètre de la loi de Poisson et l'espérance statistique du temps de séjour du modèle dans un état donné.

$$P_G(d) = e^{-l_s} \times \frac{(l_s)^d}{d!} \text{ dont les valeurs sont résumées dans le Tableau 3.1.}$$

La durée	Distribution Poissonniène	d=5s	d=10s	d=15s	Distribution gaussienne	d=5s	d=10s	d=15s
$P_G(d)$		0,0503	0,5207	0,9316		0,0010	0,4500	0,9500

Tableau 3.1. Valeurs probabilistes représentant la quantité de chance que le modèle soit dans un état pour une durée d .

Au niveau «événements» où $n=4$, le MMCHT est constitué de deux états noté $\{S_1^4, S_2^4\}$ qui sont émetteurs des événements élémentaires (Ev_1 et Ev_2) dont l'enchaînement permet de reconnaître le scénario étudié. L'événement Ev_1 correspond au fait que la personne est à proximité du guichet et l'événement Ev_2 qui correspond au fait que la personne est face au guichet et effectue une action au niveau de distribution «retirer de l'argent au distributeur».

La Figure 3.11 montre la définition des zones d'intérêts prédéfinies dans la scène.

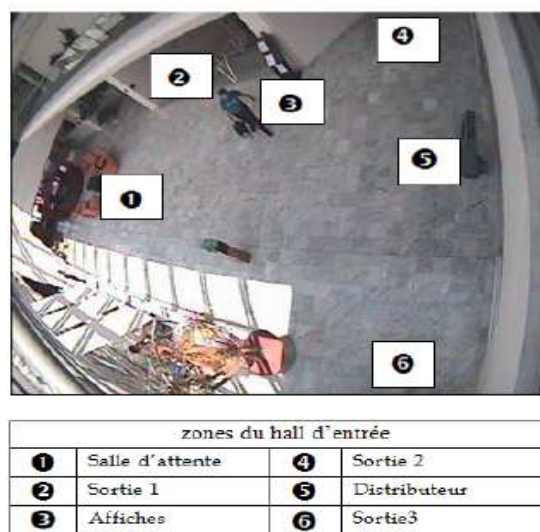


Figure 3.11. Reconnaissance d'événements «utiliser un distributeur» dans des séquences vidéo tests de la base de données PETS'04.

La Figure 3.12 illustre des résultats obtenus sur deux séquences d'images différentes PETS 2004. D'autre part, les graphes de la Figure 3.13 et la Figure 3.14 indiquent la probabilité de reconnaissance du scénario étudié en fonction du numéro de séquence selon la distribution probabiliste du temps de séjour utilisé (*poissonniène ou gaussienne*).



(a) Image d'analyse et reconnaissance de l'événement le piéton quitte la salle d'attente et va au distributeur, puis il passe un temps de 10 à 15 secondes.

(b) Image d'analyse et reconnaissance de l'événement le piéton quitte la salle d'attente et va au distributeur, puis passe un temps de 10 à 15 secondes et se dirige vers la sortie 3.

Figure 3.12. Reconnaissance d'événements «utiliser un distributeur» dans des séquences vidéo tests de la base de données PETS'04.

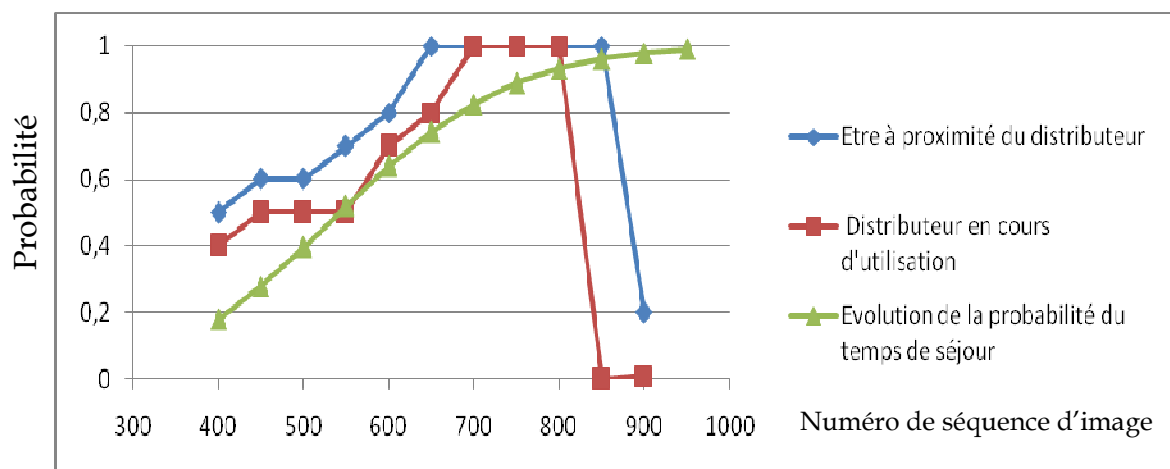


Figure 3.13. Taux de reconnaissance d'événements «utiliser un distributeur» dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution de Poisson pour la modélisation du temps de séjour.

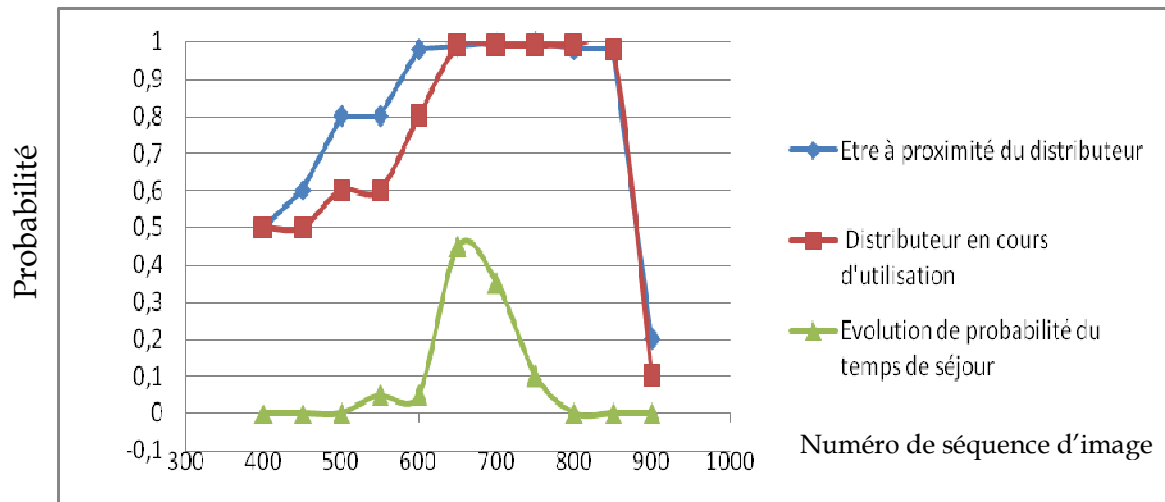


Figure 3.14. Taux de reconnaissance d'événements «utiliser un distributeur» dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution gaussienne pour la modélisation du temps de séjour.

Nous procédons par la même technique que celle exploitée pour reconnaître le scénario «utiliser un distributeur» dans le cadre de l'analyse des événements du hall d'entrée d'une gare afin de reconnaître cette fois le scénario «abandon de bagage» qui est un scénario test bien connu dans la communauté de vision par ordinateur. Pour cela nous construisons un Modèle MMCHT avec un laps de temps représenté à la Figure 3.13. Dans cet exemple, nous modélisons le fait qu'un piéton avec un bagage rentre dans le hall puis se dirige vers la zone où il dépose son bagage dans une région d'intérêt qui se traduit par la séparation d'un piéton d'un objet statique (capacité du système de reconnaissance à détecter une situation de dépôts de bagage) et enfin que le piéton s'éloigne de cet objet pendant une durée d modéliser par une distribution probabiliste. Des caractéristiques spécifiques telles que la taille du bagage, le lieu du dépôt (région d'intérêt) du bagage en fonction de la caméra sont nécessaires au système d'interprétation pour la reconnaissance de ce scénario. Les Figures 3.14 et Figure 3.15 illustrent les résultats de reconnaissance obtenus sur les différentes séquences issues des bases PETS'2004.

La structure organisationnelle hiérarchique est déclinée comme dans le cas étudié précédemment en niveau (scène, les zones d'intérêts, niveau zones-actions et niveau actions) et les éléments de base qui sont bien connues.

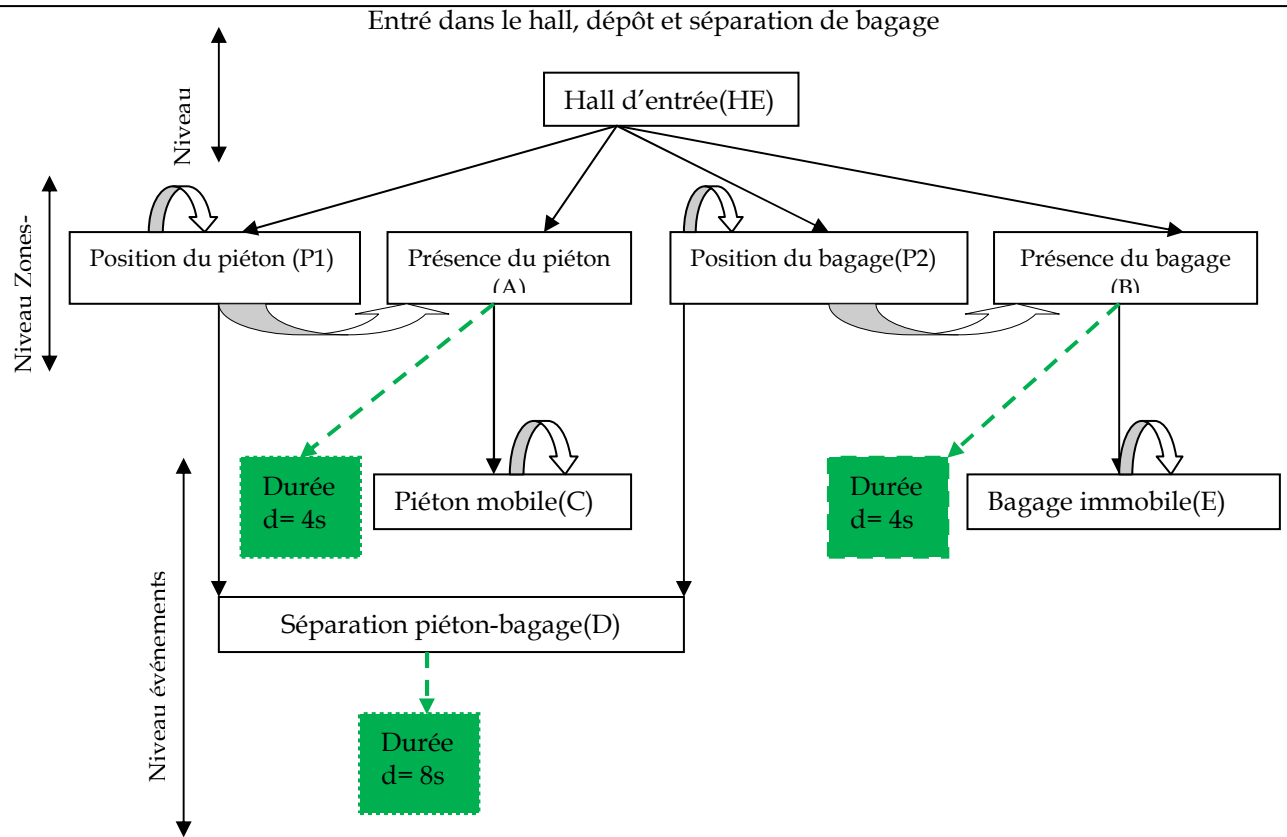


Figure 3.15. Taux de reconnaissance d'événements «*abandon de bagage*» dans des séquences vidéo tests de la base de données PETS'04 en exploitant une distribution gaussienne pour la modélisation du temps de séjour.



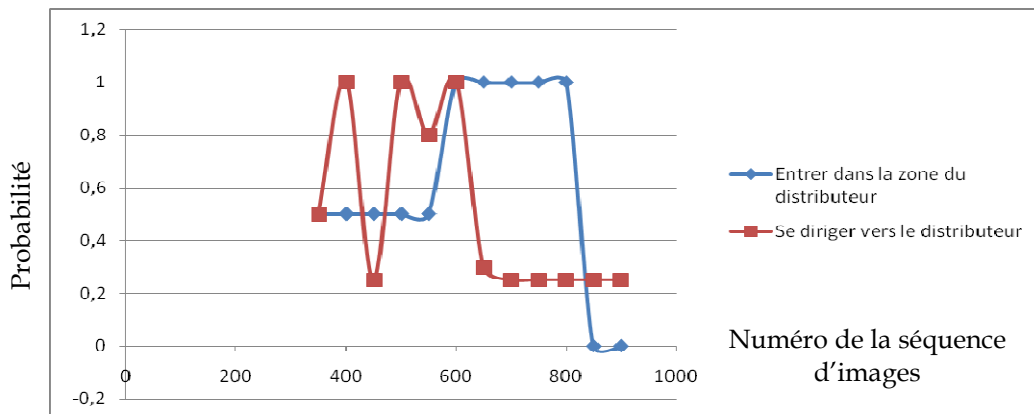
(a) Image d'analyse et reconnaissance de l'événement le piéton abandonne le bagage à la salle d'attente et quitte la salle d'attente et va au distributeur, puis il passe un temps de 10 à 15 secondes au distributeur.

(b) Image d'analyse et reconnaissance de l'événement le piéton abandonne son bagage au distributeur pendant un durée de 40s, et quitte le hall d'entrée de la gare et sort.

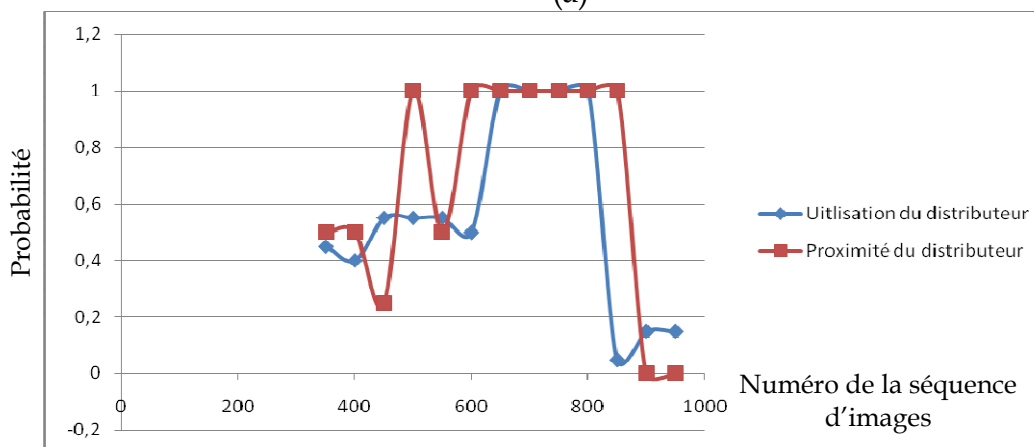
Figure 3.16. Taux de reconnaissance d'événements Reconnaissance d'événements «*bagage abandonné*» dans des séquences vidéo tests de la base de données PETS'06.

Les graphes de la Figure 3.16 montrent l'évolution de la probabilité en fonction du nombre de séquence d'image dans le cadre de la reconnaissance du scénario «*abandon de bagage*». Après une analyse de ce graphes, nous pouvons constater que :

- La Figure 3.16a illustre la reconnaissance de l'événement le piéton marche en avant, entre dans le hall puis continu à marché en avant. Cet événement modélise l'action entrée dans le hall qui est détecté entre l'image 820 et l'image 837 de la séquence où le piéton se trouve réellement à l'entrée du hall dans la vidéo.
- La Figure 3.16b illustre la reconnaissance de l'événement le piéton en arrêt, après son entrée dans le hall puis sa sortie du hall. Cet événement modélise l'action le piéton sort du hall qui est détecté entre l'image 2065 et l'image 2125 de la séquence où le piéton se trouve réellement à sortie du hall dans la vidéo.



(a)



(b)

Figure 3.17. Reconnaissance d'événements «*abandon de bagage*» les graphes (a) et (b) correspond aux résultats de la probabilité de reconnaissance en fonction du numéro de la séquence de la base test PETS'04.

3.5. Apports du modèle MMCHT

Le modèle de Markov standard, malgré sa grande utilisation avec succès dans de nombreuses applications est à améliorer face à son incapacité à modéliser des scénarios complexes. En réponse à ces lacunes, plusieurs extensions ont été proposées en intégrant des contraintes liées à la structure complexe généralement spatio-temporelle des scénarios à reconnaître. Une première catégorie d'extensions permettant de gérer la structure interne hiérarchique des scénarios et de diverses manières a été proposée et exploitée par Hung H. Bui pour l'apprentissage et la reconnaissance de séquences d'états avec des résultats très encourageants [BVW01, BVW02]. Une seconde catégorie d'extensions permettant de modéliser de façon explicite la durée de l'état du modèle a été proposée et exploitée par Thi V. Duong dans une application de reconnaissance de situations anormales dans des vidéos d'activités humaines [DH07, DVB+05]. Inspiré par la performance de ces deux grandes catégories d'extensions et surtout motivé par d'excellents résultats [BPV04, LVW+04a, LVW+04b, DVB+09] obtenus par ces auteurs dans diverses applications, nous avons proposé dans ce travail un modèle hybride qui est basé sur ces deux extensions. Notre approche a donc la capacité de gérer simultanément deux niveaux d'abstractions (hiérarchique et temporel) dans lesquels résident l'avantage majeure et notre contribution par rapport aux travaux de H. Bui et T. V. Duong. Ce modèle hybride permet d'exploiter la capacité des MMCs hiérarchiques à modéliser l'organisation spatiale des scénarios complexes et profiter de la possibilité qu'offre le modèle de semi-Markovien pour la quantification explicite de la durée des états des scénarios complexes.

Conclusion

Dans ce chapitre, nous avons étudié les mécanismes de reconnaissance séquences d'états par exploitation des MMCs et leurs extensions. Nous avons développé un modèle intégrant une structure hiérarchique pour la gestion des scénarios ayant ce type d'architecture et permettant de modéliser explicitement la durée d'état passée par le modèle, ce qui s'avère très important dans le cadre de notre application car le temps représente un paramètre intéressant pour différencier des scénarios ayant la même topologie (structure physique organisation soit les mêmes enchaînements de séquences). Les mécanismes algorithmiques pour ce modèle appelé MMCHT ont été enfin décrits tant bien pour la phase d'apprentissage que pour la phase de reconnaissance proprement dite. Nous avons dans un premier instant validé ces différentes approches à travers différentes expérimentations effectuées sur des données réelles dans la communauté de vision par ordinateur (PETS'04 et PETS'14). Plus tard, une présentation et analyse des résultats obtenus de leurs applications sur notre base de données vidéos contenant des événements importants pendant le séjour d'un patient sont présentées dans le chapitre 5 de ce mémoire. Malgré les performances enregistrées avec ces approches, la flexibilité et la robustesse de la gestion d'incertitude et d'imperfections des données reste une question d'actualité que nous nous proposons de résoudre en exploitant un cadre de raisonnement plus générique que sont les Modèles de Croyances Transférables (MCT).

Chapitre 4

Reconnaissance de scénario par les modèles de Markov Cachés Hiérarchico-Temporels et Crédibilistes

Introduction

La reconnaissance d'activités humaines est l'une des applications qui a connu beaucoup de succès ces dernières années en vision par ordinateur. Plusieurs outils mathématiques ont été développés comme solutions à ces applications. Généralement, pour ces applications, on ne dispose pas d'informations a priori sur les modèles des séquences mais seulement des observations numériques, continues ou discrètes, bruitées. Au vue de ces caractéristiques inhérentes aux données à manipuler par ces applications, il est nécessaire de développer des algorithmes capables de s'adapter ou de prendre en compte ces différentes spécificités. Des outils tels que les modèles de Markov constituent une solution assez efficace et puissante à ce problème. Cependant, quelques lacunes existent dans ces modèles qu'il convient de corriger pour rendre ces techniques plus robustes en tenant compte des imperfections des données et des modèles mis en jeux. Ainsi, après avoir fait un bref aperçu global sur les outils de gestion d'incertitude, nous introduisons une description des extensions, des algorithmes d'apprentissage des paramètres du modèle et de reconnaissance de séquences d'états dans le cas des MMCs crédibilistes afin d'utiliser les fonctions de croyances construites dans le cadre des Modèles des Croyances Transférables (MCT). L'objectif d'une telle extension est d'exploiter la généricité des fonctions de croyances et de profiter de la puissance et l'efficacité des algorithmes du MMCs. Les algorithmes proposés on été validés à travers différentes expérimentations effectuées sur des données réelles de la communauté de vision par ordinateur (la séquence PETS'14 dans un contexte multicaméras).

4.1. Outils de gestion d'incertitude type crédibiliste

L'imprécis et l'incertain peuvent être considérés comme deux points de vue antagonistes sur une même réalité qu'est l'imperfection de l'information. L'imprécis concerne le contenu de l'information tandis que l'incertain est relatif à la mauvaise connaissance de sa vérité. La théorie de l'évidence permet de combiner des informations provenant de plusieurs sources d'informations disparates pour produire un résultat d'inférence. Celui-ci peut aider à la prise de décision, avec un degré de certitude ou de confiance, pour un problème quelconque du monde réel (où l'on est constamment confronté au problème d'utilisation de plusieurs sources d'informations provenant de jugements humains statistiques, ou autres [San91]).

4.1.1. De la théorie de l'évidence aux modèles des croyances transférables

L'article initial de Dempster [DPA67] inspiré des travaux de Shafer lui ont permis d'écrire son livre [Sha76], qui est fondateur de la théorie. Après cet important ouvrage, plusieurs modifications et généralisations ont été apportées. De même, différents noms ont ensuite été utilisés pour désigner la théorie : «*théorie de Dempster-Shafer*», «*théorie de l'évidence*», et enfin «*la théorie des fonctions de croyances*».

La théorie de l'évidence ou des croyances a pour objectif d'obtenir une représentation formelle de l'aspect imprécis et incertain d'une information. On juge l'incertitude d'une information à l'aide de qualificatifs tels que «*probable*», «*possible*», «*nécessaire*», «*plausible*», «*crédible*» [DP87]. Cette théorie concerne la modélisation et la quantification de la crédibilité attribuée à des faits. Du point de vue mathématique, la théorie des fonctions de croyance étend à la fois les approches ensemblistes et probabilistes de représentation de l'incertain. Une fonction de croyance peut être vue comme un ensemble généralisé et une mesure non additive. Elle constitue une extension de notions probabilistes (conditionnement, marginalisation) et ensemblistes (intersection, union, inclusion, etc.).

Dans le modèle de base de Dempster-Shafer [Dem67, Dem68, Sha76], la règle de combinaison suppose l'indépendance des sources d'informations alors que dans de

nombreuses situations, les sources ne le sont pas car elles partagent certaines informations et peuvent être identiques disjointes ou non. Par exemple, les classifieurs peuvent être entraînés sur des bases d'apprentissage communes. Il y a nécessité de disposer de règles de combinaison tolérant la dépendance et la redondance des informations combinées.

Les modèles des croyances transférables proposés par Philips Smets [SK94] constituent une reformulation (extension) du modèle de base de la théorie de l'évidence et offrent des opérateurs prudents de combinaison en prenant en compte la nature et l'état des sources d'informations. Nous nous proposons d'utiliser le cadre des modèles de croyances transférables pour la combinaison, la gestion d'incertitude et de l'incomplétude des données dans notre système d'interprétation de séquence vidéo médicale.

4.1.2. Formalismes et fonctions de croyances du MCT

Le modèle des croyances transférables est un cadre formel générique [SK94] pour la représentation et la combinaison des connaissances. Ce cadre de travail propose des outils performants pour la gestion de l'incertitude et de l'imprécision. Le MCT est basé sur la définition de fonctions de croyances fournies par des sources d'information pouvant être complémentaires, redondantes et éventuellement non-indépendantes. Il propose un ensemble d'opérateurs très prudents permettant de combiner ces fonctions. Il est naturellement employé dans le cadre de la fusion d'informations pour améliorer l'analyse et l'interprétation de données issues de sources d'informations multiples [Ram67].

4.1.2.1. Formalisme et cadre de raisonnement

Le cadre de raisonnement dans le MCT comporte deux niveaux représentant une abstraction de la logique de ce raisonnement à savoir : le niveau crédal et le niveau pignistique.

- Le niveau «*crédal*» comprend principalement l'étape de modélisation, c'est-à-dire le modèle de représentation des connaissances sous forme de fonctions de

croyance [Den04] qui est la partie statique, ainsi que la combinaison de ces croyances appelée révision qui correspond à la partie dynamique du modèle.

- Le niveau «*pignistique*» est le siège de la prise de décision qui impose une transformation des fonctions de croyances en distributions de probabilité. Chaque masse de croyance normalisée est partagée équitablement en se basant sur le principe de la raison suffisante, qui stipule qu'en l'absence de raison de privilégier une hypothèse plutôt qu'une autre, on doit supposer les hypothèses équiprobables [Den08].

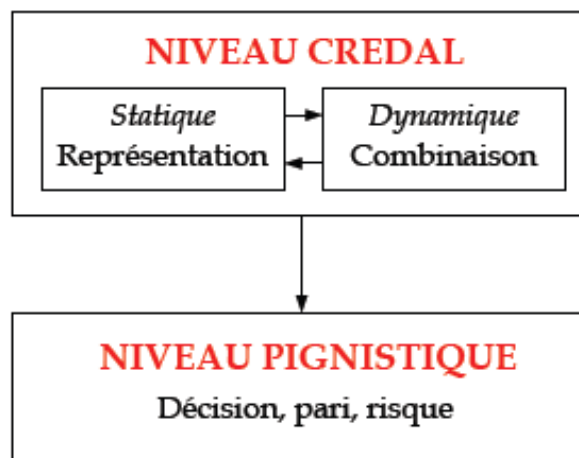


Figure 4.1. Représentation abstraite des mécanismes dans le Modèle des Croyances Transférables [Ram07].

Cette décomposition en deux niveaux différencie le MCT des autres théories et méthodes, en particulier la théorie des probabilités au sein de laquelle la représentation des connaissances, la combinaison et la décision sont effectuées sans distinguer les niveaux d'abstraction. Le MCT répond aux questions génériques suivantes :

- 1) comment représenter la connaissance d'une source d'informations sous la forme de fonctions de croyances?
- 2) comment combiner plusieurs sources d'informations représentées par les fonctions de croyance afin de résumer l'information et améliorer l'information pour une meilleure prise de décision?
- 3) comment prendre une décision à partir de fonctions de croyances?

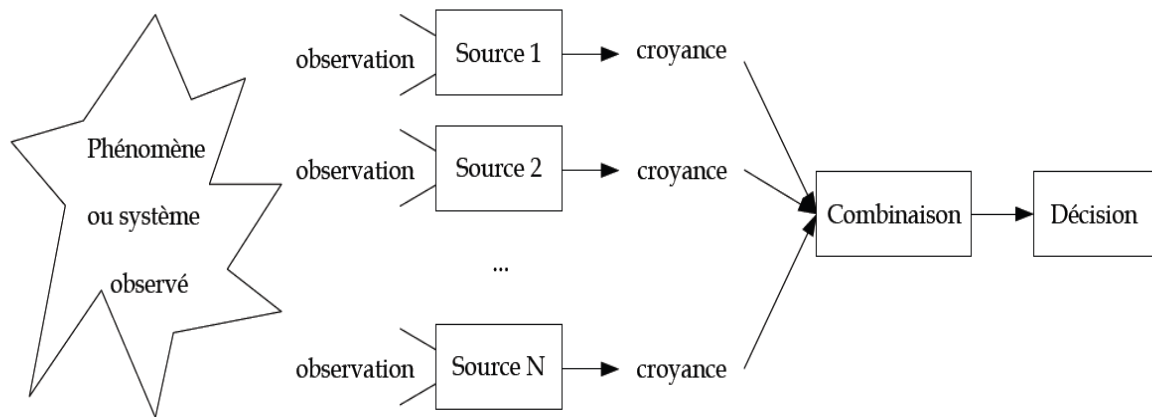


Figure 4.2. Illustration du processus de fusion d'informations multi sources [Ram07].

Au premier niveau, chaque source de croyance émet une opinion pondérée sur l'état du système observé. Au second niveau, les différentes sources sont combinées afin de synthétiser la compréhension du système et enfin au troisième niveau on améliore la prise de décision sur l'état du système après combinaison.

4.1.2.2. Définitions et fonctions de croyances

La théorie des croyances utilise les fonctions de masses m , de plausibilité pl , de crédibilité Cr et communalité q pour représenter à la fois l'incertitude et l'imprécision relatives à une information provenant de plusieurs sources. Les fonctions de masses sont définies sur tous les sous-ensembles de l'espace de discernement Ω (contenant les classes intéressantes) et pas simplement sur les singletons comme les probabilités qui ne mesurent que la probabilité d'appartenance à une classe de données. Le cadre de discernement contient toutes les hypothèses H_i nécessaires à la description complète d'une situation présentée. Cet espace de discernement s'écrit : $\Omega = \{H_1, H_2, H_3, \dots, H_K\}$. L'indice K indique le nombre total d'hypothèses. Pour modéliser et quantifier la crédibilité accordée à des faits, dont on ne connaît pas la probabilité d'occurrence, on utilise des fonctions de croyance qui indiquent l'ordre de confiance attribuée à ces faits. L'univers de référence Ω selon la théorie de Dempster-Shafer est supposé fini et détermine les coefficients de croyance où les degrés de confiance. Ceux-ci sont obtenus en répartissant une masse globale de croyance égale à 1 entre tous les événements possibles et en attribuant à chacun le degré $m(B)$ avec lequel le groupe d'observateurs

croit en sa réalisation. Dans la théorie de Dempster-Shafer, le raisonnement porte sur des parties notées $2^\Omega = P(\Omega)$ qui représentent l'ensemble de tous les sous-ensembles de Ω . Un sous-ensemble arbitraire est noté B_i et par conséquent l'ensemble des parties notées 2^Ω peut être représenté par les sous ensembles B_i par l'équation 4.1 :

$$P(\Omega) = \{B_1, B_2, \dots, B_{2^k}\} \quad (4.1)$$

Un sous ensemble qui en l'union de plusieurs hypothèses est appelé une proposition par Shafer [Sha76]. Si un sous-ensemble consiste uniquement en une hypothèse H_i , il est appelé un singleton. Nous présentons ici une forme simplifiée de la théorie des croyances tirée de Smets, et de Dubois et Prade [SMD⁺88] et [Bouc95].

Une masse de croyance est une fonction m^Ω qui attribue un coefficient compris entre 0 et 1 aux parties de Ω de telle sorte que :

$$\sum_{B \in P(\Omega)} m^\Omega(B) = 1 \quad \text{et} \quad m^\Omega(\emptyset) = 0 \quad (4.2)$$

La fonction de croyance de masse notée m^Ω est définie par :

$$\begin{aligned} m^\Omega : 2^\Omega &\rightarrow [0, 1] \\ B &\mapsto m^\Omega(B) \end{aligned} \quad (4.3)$$

La quantité $m^\Omega(B)$ représente la croyance exacte dans l'événement représenté par B. Si $m^\Omega(B)=1$ et $\forall C \neq B, C \in P(\Omega), m^\Omega(C)=0$, par conséquent l'événement B est certain dans le sens que l'un des éléments de B est la valeur cherchée. Toutefois, on ne sait pas forcément de quel élément de B, il s'agit, sauf si B est un singleton. D'autre part, si $B = \Omega$, alors l'information $m^\Omega(B)=1$ ne nous apprend rien, sinon que nous sommes dans une situation d'ignorance totale.

On appelle élément focal toute partie non vide E de Ω telle que $m(E) \neq 0$ et on note ε leur ensemble.

On appelle corps d'évidence le couple (ε, m) . A partir de la fonction de masse, de nombreuses autres fonctions ont été définies pour des utilisations spécifiques à l'application et selon le contexte et la nature des données. A partir de la fonction de masse définie plus haut dans le cadre de la théorie de l'évidence développée par Dempster-Shafer, d'autres fonctions en relation biunivoque peuvent être déduites et

représentent la même information mais sous une forme différente dans le cadre du MCT. Ces fonctions ont diverses interprétations et sont utilisées essentiellement pour simplifier les calculs de combinaisons selon le contexte de l'application. Parmi ces fonctions nous pouvons citer les plus utilisées telles que : la plausibilité (pl), la Crédibilité ou Croyance (Bel ou Cr), Implacabilité (b), et la Communalité (q) sont définies de 2 vers $[0, 1]$ avec la somme des éléments focaux non forcément unitaires. Dans la suite de ce document, nous faisons une présentation de ces fonctions avec leur utilité [Ram07] en redéfinissant la fonction masse de croyance par sa distribution nommée BBA (Basic Belief Assignment) dans le cadre du MCT. Dans le MCT, une distribution de masse de croyance notée BBA fournie par une source S est un ensemble de masses de croyance concernant des propositions $B \subseteq \Omega$ (ou de manière équivalente $B \in 2^\Omega$ vérifiant l'équation 4.4:

$$\sum_{B \subseteq \Omega} m_S^\Omega(B) = 1 \quad (4.4)$$

S est appelée la source de croyance qui est souvent inscrite en indice de la fonction de masse, on note par exemple m_S^Ω . La masse $m_S^\Omega(B)$ représente la part de croyance de la source S . Il existe des BBA particuliers qui sont entre autres la BBA normale $\{m^\Omega(\phi) = 0\}$, la BBA vide $\{m^\Omega(\phi) = 1\}$, la BBA dogmatique, la BBA simple, La BBA catégorique, la BBA consonante et la BBA sans conflit interne. Une BBA normale est obtenue après une opération de normalisation qui est définie par l'équation 4.5 :

$$m_*^\Omega(B) = \frac{m^\Omega(B)}{1 - m^\Omega(\phi)} \quad \forall B \subseteq \Omega, B \neq \phi \quad \text{avec} \quad m_*^\Omega(\phi) = 0 \quad (4.5)$$

- La *plausibilité* notée pl d'un élément B est la part maximale de croyance qui pourrait soutenir B :

$$pl^\Omega(B) = \sum_{C \cap B \neq \phi} m^\Omega(C), \quad \forall B \subseteq \Omega \quad (4.6)$$

- La *crédibilité* (ou *croyance*) notée bel d'un élément B est la part de croyance spécifiquement (sans le conflit) allouée à B :

$$bel^\Omega(B) = \sum_{\phi \neq C \subseteq B} m^\Omega(C), \quad \forall B \subseteq \Omega \quad (4.7)$$

- L'*implacabilité* notée b d'un élément B est la somme des masses allouées aux sous-ensembles de B tels que leur véracité implique la véracité de B :

$$b^\Omega(B) = \sum_{C \subseteq B} m^\Omega(C) = bel^\Omega(B) + m^\Omega(\phi), \quad \forall B \subseteq \Omega \quad (4.8)$$

- La *communalité* notée q d'un élément B représente la somme des masses allouées aux sur-ensembles de B et qui ont B en commun :

$$q^\Omega(B) = \sum_{C \supseteq B} m^\Omega(C), \quad \forall B \subseteq \Omega \quad (4.9)$$

Chacune de ces fonctions de par son utilité est adaptée à des applications spécifiques. Elle est obtenue par transformation de la fonction masse en ces différentes fonctions. La fonction masse et la crédibilité permet de représenter la connaissance. La plausibilité est utilisée dans le cadre du théorème de Bayes généralisé et du maximum de vraisemblance. Il existe principalement deux types de règles de combinaison d'informations ou de fusion d'informations. La première est celle conjonctive où la fonction communalité est utilisée et la seconde est celle disjonctive où la fonction implacabilité est exploitée.

Le Tableau 4.1 donne une indication sur l'utilisation de ces fonctions dans le cadre de ces travaux. Un résumé de ces transformés est donné dans [Sme02a] et il existe des outils comme TBMLab³ qui contiennent quelques procédures Matlab pour les différentes conversions.

<i>Sigle</i>	<i>Nom</i>	<i>Fonction ou utilité</i>
<i>M</i>	<i>Masse</i>	<i>Représentation de la connaissance</i>
<i>Pl</i>	<i>Plausibilité (Eq. 4.3)</i>	<i>Vraisemblance ou TBG</i>
<i>Bel</i>	<i>Crédibilité (Eq 4.7)</i>	<i>Représentation de la connaissance</i>
<i>B</i>	<i>Implacabilité (Eq. 4.8)</i>	<i>Règle de combinaison disjonctive (DRC)</i>
<i>q</i>	<i>Communalité (Eq. 4.9)</i>	<i>Règle de combinaison conjonctive (CRC)</i>

Tableau 4.1. Les fonctions de croyance et leur utilité.

Le Tableau 4.2 présente un exemple de résultat de la transformation de la fonction masse m donnée de la colonne deux en fonctions de croyances de *plausibilité* (pl), *crédibilité* (Cr), *implacabilité* (b), et *communalité* (q).

	m	pl	Bel	b	q
ϕ	0.1	0	0	0.1	1
$\{w_1\}$	0.07	0.46	0.08	0.17	0.46
$\{w_2\}$	0.12	0.44	0.13	0.22	0.44
$\{w_3\}$	0.31	0.49	0.34	0.41	0.49
$\{w_1, w_2\}$	0.22	0.59	0.45	0.51	0.31
$\{w_1, w_3\}$	0.08	0.78	0.51	0.56	0.17
$\{w_2, w_3\}$	0.01	0.83	0.48	0.54	0.1
$\{w_1, w_2, w_3\}$	0.09	0.9	1	1	0.09

Tableau 4.2. Résultat de la transformation d'une fonction masse m donnée en *plausibilité* (pl), *crédibilité* (Cr), *implacabilité* (b), et *communalité* (q).

4.2. Opérations de bases dans le cadre du MCT

Plusieurs opérations peuvent être effectuées avec les fonctions de croyance pour traiter des problèmes de modélisation et de représentation des connaissances, de combinaisons complexes d'informations issues de sources disparates et de prise de décision. Nous pouvons classer en deux groupes ces opérations : celles qui sont basiques, bien connues et celles que nous appelons opérations avancées [Ram09] qui concernent l'utilisation du *Théorème de Bayes Généralisé (TBG)*, combinaison d'informations issues de différents cadres de discernement, le conditionnement et le déconditionnement des masses pour la fusion de fonctions de croyances conditionnelles, le *Principe de Minimum d'Information généralisée (PMI)*. Le choix d'une masse de croyance parmi un ensemble de masses possibles, est motivé par l'estimation de la quantité d'informations qu'elle transporte, c'est-à-dire, qu'il faut qu'elle soit *la moins informative possible* autrement dit qu'elle soit *la moins engagée*. C'est le *principe de minimum d'information*. Le *Framework* qu'est le **MCT** offre des solutions à des

applications de diagnostic de système ou de prédiction. Nous présenterons dans ce document quelques opérations de base qui concernent la combinaison et la prise de décision.

4.2.1. Le PMI : principe du minimum d'information

De nombreuses difficultés sont à franchir lors de la phase d'estimation des distributions des fonctions de masses. Pour combiner efficacement les informations et disposer de données plus pertinentes pour la phase de prise de décision, il faudra transformer l'information mesurée à travers une distribution probabiliste en une distribution crédibiliste ou en une masse de croyance (appelé BBA : Basic Belief Assignment par E. Ramasso). La transformation (fonctions probabilistes \rightarrow fonctions de croyances) est nécessaire car les combinaisons CRC et DRC n'ont aucun intérêt lorsque les BBA combinées sont bayésiennes. Par exemple, lorsqu'on dispose d'une distribution de probabilités pignistiques $BetP$, le problème est de déterminer une BBA m dont la transformation pignistique redonne $BetP$ [Ram09]. Deux formes de solutions selon l'approche (*quantitative* ou *qualitative*) adoptée s'offrent pour résoudre ce problème.

L'*approche quantitative* consiste à calculer le taux d'incertitude des masses et de choisir celle qui maximise l'incertitude et minimisant ainsi le parti pris. Une mesure non spécifiée est basée sur la cardinalité des éléments focaux pondérés par la valeur des masses qui est estimée par l'équation 4.10 :

$$N(m) = \sum_{\phi=A \subseteq \Omega} m(A) \times \log_2(|A|) \quad (4.10)$$

Cette mesure est bornée sur $[0, \log_2(|\Omega|)]$. S'agissant d'applications exploitant cette approche quantitative, G.J. Klir, M.J. Wierman. et D. Harmanec ont proposé des techniques de mesure du taux d'incertitude relatif à une information imprécise [KW99, Har99]. De même, T. Denoeux et M.H. Masson ont développé une technique pour une application de clustering [DM04].

Une *approche qualitative* a été proposée dans [DP87a, Yag87, Sme00] et consiste à utiliser des relations d'ordre entre les fonctions de croyance avec le PMI basée

sur la non-spécificité et maximise donc la valeur de ce critère qui joue un rôle similaire au maximum d'entropie en théorie de l'information.

4.2.2. Règles de base de combinaison d'informations

La combinaison ou fusion d'informations est un traitement a priori qui permet d'améliorer la donnée qui sert à la prise de décision. Ce traitement vise à combiner des informations hétérogènes issues de plusieurs sources d'informations (distinctes ou non/fiables ou non) afin d'aider à la prise de décision. Par exemple, le cas de données imparfaites (incertaines, imprécises et incomplètes), la fusion d'informations dans un contexte crédibiliste est une solution intéressante qui consiste à combiner des informations hétérogènes issues de plusieurs sources indépendantes ou non pour une meilleure prise de décision [Blo05, DP87c, DP87d]. Plusieurs critères et opérateurs prudents de fusion de BBA ont été développés dans le cadre du MCT. Deux niveaux d'opérations peuvent être distingués : les opérations de combinaison de base et les opérations de combinaison avancées. Quelle que soit la catégorie d'opérateurs de combinaison d'informations, deux hypothèses généralement sont faites sur les sources d'informations: elles doivent appartenir au même cadre de raisonnement (évidentielles ou probabilistes par exemple) et doivent être indépendantes ou distinctes [Smet93, Smet98]. Nous présentons dans la suite de ce document quelques mécanismes de base pour la fusion d'informations basées sur les principaux opérateurs de base de combinaison tels que la somme conjonctive et la somme disjonctive avec les hypothèses de fiabilité de la source (sources fiables ou non) et de l'indépendance en probabilité-statistique qui est relative à la distinction des sources (sources distinctes ou non) [YSM01, YSM02a, YSM02b].

4.2.2.1. Combinaison d'informations issues de sources distinctes et fiables

Pour combiner des informations issues de sources distinctes et fiables dans le cadre des modèles de croyances transférables, il est proposé l'utilisation d'une règle de combinaison conjonctive (*CRC : Conjunctive Rules of Combination*) et est notée \odot . La règle de combinaison conjonctive possède plusieurs propriétés telles que la

commutativité et l'associativité mais pas l'idempotence, la linéarité (peut être décrite sous forme de masse). Le CRC permet de mettre en évidence et de gérer le conflit car lorsque l'intersection de deux éléments est vide, les sources sont généralement conflictuelles et il est nécessaire d'identifier les raisons du conflit puisque l'ensemble vide est absorbant par l'intersection. La masse ou la fonction de croyance issue d'une combinaison de type CRC peut être sous normale et pour cela, il faut une normalisation qui ramène le CRC à une règle orthogonale comme dans la théorie de l'évidence de base de Dempster-Shafer. Quelques extensions de la CRC ont été proposées pour qu'elle soit plus prudente notamment par E. Lefevre et al. [LCV02] où les fonctions de croyances sont redistribuées pour annuler l'effet de l'intersection vide qui est source de conflit lors de la combinaison d'information par T. Denoeux [Den08], où il propose un opérateur associatif, commutatif et idempotent.

En outre, la combinaison d'informations certaines nécessite une opération de conditionnement qui est définie par la règle de combinaison conjonctive. L'opération de conditionnement est particulièrement adaptée dans le cas où l'on dispose d'une information certaine mais qui peut être imprécise car elle revient à effectuer une combinaison conjonctive avec une masse catégorique sur B ($m^\Omega(B)=1$). En effet, le conditionnement d'une masse m^Ω par un élément $B \subseteq \Omega$ consiste à restreindre le cadre des propositions possibles 2^Ω à celles ayant une intersection non nulle avec B . Par conséquent, lors d'un conditionnement, les masses affectées à $C \subseteq \Omega$ sont transférées sur $C \cap B$.

La masse issue du conditionnement est parfois sous-normale ce qui nécessite une gestion du conflit présentée à la section 4.2.4. Elle est notée $m^\Omega[B](C)$ et s'effectue par l'équation 4.11 :

$$m^\Omega[B](C) = \sum m^\Omega(C \cup D), \quad \forall C \subseteq B \quad (4.11)$$

L'intersection intervenant dans la CRC produit un effet de spécialisation [Sme02a], en transférant la masse sur des sous-ensembles de cardinalité plus faible. Ce qui est un avantage très intéressant dans l'utilisation des fonctions de masse transformées en fonction de communalité car la CRC est réduite à un produit membre à membre qui simplifie de ce fait les calculs.

4.2.2.2. Combinaison d'informations issues de sources distinctes et non fiables

La combinaison ou la fusion des informations est une opération qui nécessite une importante attitude de prudence pour ne pas affecter la qualité de la donnée résultant de l'opération de fusion d'informations sur laquelle doit se baser la prise de décision. Il importe de définir des critères de fusions très prudents qui tiennent compte de l'état de la source d'informations qui représente un facteur affectant la qualité du résultat de la combinaison. Pour prendre en compte la fiabilité ou non des sources d'informations en combinaison, le Framework MCT propose une règle de combinaison disjonctive dite DRC : Disjunctive rule of combination [DP86b, Sme93] et est noté : \cup .

La DRC permet de combiner des informations issues d'un ensemble de sources dont l'une est au moins fiable même si l'on ne dispose d'aucune connaissance sur la source pour quantifier sa fiabilité. S'agissant des propriétés de cette règle, nous pouvons dire qu'elle est associative, commutative et non idempotente comme la CRC. Cette règle est conservatrice car elle permet de transférer les masses sur des sur-ensembles d'éléments focaux et par conséquent produit un effet de généralisation a contrario de la CRC qui produit un effet de spécialisation [Sme02]. Pour faciliter les opérations de calcul, la fonction implacabilité est un bon compromis qui permet de réduire la DRC à un produit membre à membre avec une opération linéaire décrite sous forme matricielle. Le caractère parfois trop conservateur de la DRC peut être un handicap en transformant une masse de croyance vide complètement non-informative lorsque les sources ne sont pas fiables. Pour ce faire, il est opportun de prendre en compte la fiabilité de la source lorsqu'il est possible de la quantifier afin d'appliquer une règle de combinaison moins conservatrice que la DRC. La règle utilisée dans ce cas est nommée règle d'affaiblissement simple. Elle a été introduite par Shafer [Sha76], axiomisée par Smets [Sme93] et généralisée par Mercier et Denoeux [MQD07, Mer06] et est nommée dans ce dernier cas *affaiblissement contextuel*.

L'affaiblissement d'une masse m^Ω est défini comme la pondération de chaque masse $m^\Omega(B)$ de la distribution par un coefficient de fiabilité qui vaut $1 - \alpha$ où $\alpha \in [0, 1]$

et représente le taux d'affaiblissement. Elle s'obtient en utilisant l'équation 4.12a et 4.12b:

$$m^\Omega(B) = (1 - \alpha) \times m^\Omega(B), \quad \forall B \subset \Omega \quad (4.12a)$$

$$m^\Omega(\Omega) = (1 - \alpha) \times m^\Omega(\Omega) + \alpha \quad (4.12b)$$

4.2.3. Prise de décision à partir de fonctions de croyance

Une importante étape, pour un système de fusion d'informations quel que soit l'outil mathématique servant de cadre de gestion d'incertitude, est celle de la prise de décision. Décider, c'est choisir une hypothèse sur un cadre de pari, appelé généralement le cadre de discernement et noté Ω . Cette décision peut être prise automatiquement par le système lorsque les critères de prise de décisions sont bien définis et connus du système. Cependant, il existe des applications sensibles qui nécessitent une phase de validation de la prise de décision qui en conséquence est laissée à la charge d'un opérateur ou un utilisateur. C'est le cas par exemple des applications d'aide au diagnostic dans le domaine médical. Dans le cadre des modèles de croyances transférables, la prise de décision s'appuie sur deux fonctions : la probabilité pignistique et la plausibilité.

4.2.3.1. Prise de décision à partir de la probabilité pignistique

La distribution de probabilités pignistiques [Sme05] est très utilisée à la phase de prise de décision dans le MCT. La transformée pignistique consiste à répartir de manière équiprobable la masse d'une proposition B sur les hypothèses contenues dans cette proposition. Autrement, elle consiste à partager équitablement chaque masse de croyance si au besoin normalisée en se basant sur le principe de la raison suffisante, qui stipule qu'en absence de raison, de privilégier une hypothèse plutôt qu'une autre. On doit ainsi, supposer les hypothèses équiprobables (Den07). On obtient ainsi, une distribution de probabilité pignistique notée $BetP\{m^\Omega\}$. Formellement, elle est définie par l'équation 4.13 mais la décision est généralement opérée en choisissant l'élément possédant la plus grande probabilité pignistique:

$$\begin{aligned} \text{BetP}\{m^\Omega\} : \Omega &\rightarrow [0, 1] \\ w_k &\mapsto \text{BetP}\{m^\Omega\}(w_k) \end{aligned} \quad (4.13)$$

Où $\text{BetP}\{m^\Omega\}$ est donnée par l'équation 4.14a :

$$\text{BetP}\{m^\Omega\}(w_k) = \frac{1}{1 - m^\Omega(\phi)} \sum_{B \subseteq \Omega, w_k \in B} \frac{m^\Omega(B)}{|B|} \quad (4.14a)$$

$$w_0 = \arg \max \text{BetP}\{m^\Omega\}(w_k), \quad \forall w_k \in \Omega \quad (4.14b)$$

Lorsqu'on utilise la transformée pignistique pour la prise de décision, on adopte un comportement rationnel en maximisant l'utilité espérée tandis que l'utilisation des fonctions de plausibilité ou de crédibilité lors de cette phase peut constituer une alternative pour adopter plutôt respectivement un comportement optimiste ou pessimiste.

4.2.3.2. Prise de décision à partir de la fonction plausibilité

La fonction de plausibilité permet de se mettre dans une attitude plutôt optimiste lors de la phase de prise de décision dans le sens où la plausibilité d'un élément représente la valeur maximale de la masse de croyance sur cet élément qui pourrait être atteinte si des informations supplémentaires parviennent au système de fusion. Dans ce cas, la phase de combinaison des informations exploite une règle conjonctive pour l'intégration de la nouvelle information. La distribution de probabilité sur les singletons à partir de laquelle une décision peut être prise est donnée par l'équation 4.15 :

$$\text{PIP}\{m^\Omega\}(w_k) = \frac{pl^\Omega(w_k)}{\sum_{w_k \in \Omega} pl^\Omega(w_k)} \quad \forall w_k \in \Omega \quad (4.15)$$

4.2.4. Gestion du conflit

Diverses techniques sont exploitées pour gérer le conflit généré lors de la phase de combinaison d'informations. En effet, l'intersection entre deux éléments focaux peut être vide lorsqu'on exploite une règle de combinaison conjonctive et par conséquent la fonction croyance résultante peut être une masse non nulle. Cette valeur non nulle est

appelée le conflit et représente la discordance entre les sources d'informations. Il est important d'analyser l'origine de ce conflit qui peut être multiple [Sme05] et parmi ces origines, nous pouvons noter la non exhaustivité du cadre de discernement, le non respect des hypothèses sous-jacentes lors de l'utilisation de la règle de combinaison conjonctive.

Une autre source possible du conflit provient de la nature de la combinaison conjonctive exploitée et de celle de la fonction de croyance manipulée (une masse par exemple). Par exemple dans ce dernier cas de génération du conflit lors de la phase de combinaison des informations, nous pouvons avoir un conflit de type «*conflit interne*» qui est issu de la combinaison d'une masse avec elle même. A cet ensemble de causes pouvant générer un conflit pendant la phase de fusion d'informations, il existe plusieurs techniques de gestion du conflit. T. Denoeux propose une règle prudente dans [Den08] pour combiner les informations avec la CRC en y intégrant la gestion du conflit généré.

Une autre idée dans le cadre de la gestion du conflit interne consiste à construire des distributions de masses consonantes en s'inspirant des sous-ensembles flous comme dans [HCR05, GCB⁺05, RPR⁺07]. Il est aussi possible de changer la règle de combinaison conjonctive et l'utiliser comme proposée plus haut comme une règle simple qui minimise le conflit [Cat03]. Cependant, le conflit n'a pas qu'un sens péjoratif, mais il peut être très utile et interprété pour répondre à d'importantes questions. La quantité de masse de conflit peut être interprétée comme la croyance sur une nouvelle hypothèse dans le cas d'un modèle de croyance transférable en monde ouvert [DS04]. Il peut aussi être utile pour regrouper des fonctions de croyance [SCh04, DM04], associer à des objets [AS01], détecter de nouveaux objets [DS04] ou suivre des cibles [RS06].

4.3. Mécanismes pour la reconnaissance de scénario par exploitation des modèles de Markov cachés hiérarchico-temporels crédibilistes

Nous présentons et exploitons dans cette section, une approche de reconnaissance de scénario basée sur les modèles de Markov cachés crédibilistes tout en

y intégrant une gestion de contrainte temporelle et de la structure hiérarchique des scénarios. Ce dernier est un cadre de raisonnement plus générique que la théorie de probabilité pour gérer les imperfections des données manipulées. Les apports du modèle sont présentés à la fin de la section afin de justifier l'opportunité de l'extension des MMC probabilistes aux MCC crédibilistes. Les premiers travaux sur l'extension des MMC à la théorie de l'évidence [Dem67, Sha76] ont été proposés par Pieczynski et al. [Pie07] pour des applications de segmentation d'images en traitements d'images satellitaires.

Des travaux plus récents de E. Ramasso dont les bases ont été posées depuis 2007 dans sa thèse [Ram07] lui ont permis de se placer dans le cadre des modèles de croyances transférables développées par P. Smets pour proposer une extension des MMCs probabilistes aux MMCs crédibilistes. Inspiré par ces travaux, nous nous proposons d'intégrer d'une part à ce modèle de MMCs Crédibilistes un paramètre de type probabiliste de gestion du temps qui est par suite fusionné dans une masse puis transformé en une fonction de communalité. D'autre part, une topologie de type arborescente s'avère aussi importante pour gérer la structure arborescente des événements à reconnaître par le système. Le cadre de reformulation des MMCs permet de combiner de façon élégante et robuste avec des opérateurs prudents des informations et la gestion efficace du conflit lors de la fusion et la prise en compte du doute et de l'ignorance liée aux scénarios à reconnaître.

La puissance de ce modèle crédibiliste réside dans les mécanismes de propagation crédibiliste *forward* et *backward* pour la phase de reconnaissance et l'algorithme de *Baum-Welch crédibiliste* ou celle de *Viterbi* dans le contexte de MCT pour la phase d'apprentissage des paramètres du modèle. Dans les paragraphes suivants, nous décrivons le modèle des séquences et celui des observations, les mécanismes de base pour l'apprentissage et la reconnaissance dans le cadre du MCT.

4.3.1. Modèles de Markov cachés Crédibilistes

Les modèles de Markov cachés crédibilistes que nous présentons dans cette section représentent une extension des modèles de Markov cachés probabilistes {section 3.1} qui ont été présentés dans le chapitre 3. De même, tous les mécanismes présentés au

niveau des MMCs probabilistes restent valables pour ceux crédibilistes avec des variables sous forme de croyance. Le modèle est composé d'un modèle d'observations et d'un modèle de séquences d'états dit cachés générés par les observations.

- Un MMC crédibiliste (MMCCr) nommé λ_{Cr} est défini par un ensemble d'états noté : $\Omega_t = \{S_1^t, S_2^t, \dots, S_K^t\}$ et une distribution de fonctions de croyance peut être une masse évidentielle (m) ou une autre transformation entre la plausibilité (pl), ou les communalités (q).
- Dans notre modèle, nous proposons d'utiliser les communalités car elles sont beaucoup plus adaptées à notre contexte d'application où la règle de combinaison utilisée pour fusionner les informations est une règle conjonctive. Les sources d'informations étant distinctes et fiables (par exemple le degré de confiance accordé à une caméra pour sa qualité de détection et de suivi d'objets en mouvement dans la scène). D'autre part, l'utilisation des communalités rend les calculs moins lourds et fastidieux tout en fournissant des expressions proches de celles probabilistes. En conséquence, ceci est un avantage majeur en complexité algorithmique. Le modèle d'observation noté O est constitué d'alphabets d'observations distinctes selon M symboles tel que : $O = \{O_1, O_2, O_3, \dots, O_M\}$.
- A_{Cr} est la fonction de croyance équivalente à la probabilité de transition entre les états : $A_{Cr} = \{q_{a_i, j}^{\Omega_t}\}$ où $q_{a_i, j}^{\Omega_t}$ est obtenue par une transformation de la probabilité $P(S_{t+1} = s_j | S_t = s_i)$ en fonction de croyance de type communalité (q);
- B_{Cr} est la fonction de croyance équivalente à la probabilité d'observations (probabilité d'émission, croyance) : $B_{Cr} = \{q_{b_j}^{\Omega_t}(m)\}$ où $q_{b_j}^{\Omega_t}(m)$ est obtenue par une transformation de la probabilité $P(O_t = O_m | S_t = S_j)$ en fonction de croyance de type communalité (q);
- q^{Ω_0} est la probabilité d'état initial : $q^{\Omega_0} = \{q_i^{\Omega_0}\}$ $\Pi = \{\pi_i\}$ où $q_i^{\Omega_0}$ est obtenue par une transformation de la probabilité $P(S_1 = S_i)$ en fonction de croyance de type communalité (q).

Dans la suite du document, nous utiliserons comme présentée dans le Tableau 4.3, la dénomination des variables crédibilistes correspondantes aux variables probabilistes.

Nom des variables et algorithmes	Probabiliste	Crédibiliste
Variable α (<i>forward</i>)	α	$q_{\alpha}^{\Omega_t}$
Variable β (<i>backward</i>)	β	$q_{\beta}^{\Omega_t}$
Variable γ (<i>Baum-Welch</i>)	γ	$q_{\gamma}^{\Omega_t}$
Variable ξ (<i>Baum-Welch</i>)	ξ	$q_{\xi}^{\Omega_t}$

Tableau 4.3. Présentation de la correspondance entre notations probabilistes et celles crédibilistes.

4.3.2. Gestion de la temporalité des séquences et de la hiérarchie des scénarios par le modèle des MMCs crédibilistes

La gestion de l'évolution temporelle des séquences d'états se fait par l'estimation d'une variable probabiliste et selon les distributions, utilisable comme présentée au chapitre 3. Une fois l'estimation de ce paramètre temporel effectuée, il est transformé en une fonction masse de même que le degré de confiance (probabiliste transformé en masse évidentielle) liée à la caméra qui a fourni les mesures (observations) et la probabilité de transition de l'état. Le contenu de l'ensemble de ces trois variables fusionnées dans la même et transformé en fonction de communalité. S'agissant de la gestion de l'hiérarchie, elle s'avère nécessaire et importante afin d'éviter d'utiliser autant de MMC pour la reconnaissance de scénarios ou événements à reconnaître. Alors qu'un MMC hiérarchique permettrait de gérer efficacement plusieurs scénarios ayant différentes structures organisationnelles mais se situant tous dans la même catégorie. Les mécanismes dans ce cas d'espèce ont été décrits au chapitre 3.

La gestion de la structure hiérarchique et organisationnelle des scénarios est basée sur le modèle d'affectation proposé par D. Mercier [Mer06], dans ces travaux de thèse sur la fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance [Mer06]. Son approche est une forme purement hiérarchique de l'affectation dans un cadre de décision multi-

niveaux. Une hiérarchie est un arbre possédant une racine associée à Ω dans le cadre des croyances, un niveau (composé d'un ensemble d'éléments) associé à une partition de Ω , une feuille associée à un élément singleton. Par convention, le premier niveau est le niveau composé des singletons de Ω et le dernier niveau est celui composé de Ω .

Soit $n \in \{1, \dots, N\}$, n le niveau de la hiérarchie et N le nombre de niveaux de la hiérarchie. On définit ainsi par $\Omega^{(n)}$ l'ensemble des décisions de niveau n (par exemple $\Omega^{(1)} = \Omega$ et $\Omega^{(p)} = \{\Omega\}$). $K^{(n)}$ avec $k \in \{1, \dots, K^{(n)}\}$ dénote le nombre d'éléments de $\Omega^{(n)}$ (w_k^n désigne le $k^{\text{ième}}$ élément de $\Omega^{(n)}$). La Figure 4.3 illustre un exemple de structure hiérarchique à trois niveaux.

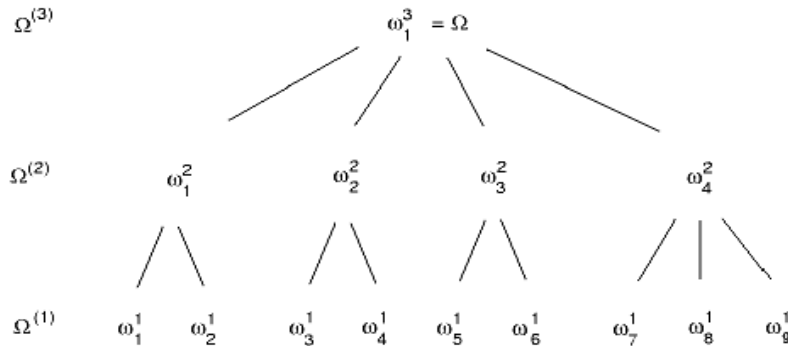


Figure 4.3. Illustration d'un processus d'hiérarchisation à trois niveaux [Mer06].

L'*algorithme 1* détaille les principales lignes d'instructions d'une procédure de *backward* de prise de décision à un niveau p de la chaîne de gestion de la structure hiérarchique :

Algorithme 1 : Procédure de prise de décision à un niveau p

- 1: Soit un **scénario** S de niveau n
 - 2: **Si** le niveau n est égal au niveau p :
 - 3: Alors, on décide que S est reconnu à un niveau p :
 - 4: **Si non** p correspond au niveau du premier élément parent de la décision contenant la vérité
 - 5: **Fin Si**
 - 6: **Si** S est **correct** au niveau p
 - 7: S est directement reconnu
 - 8: **Fin Si**
-

Nota : $p \in [n, N]$ et le nombre de décisions d de niveau n correctes au niveau p vaut $d^{n,p}$.

4.3.3. Apprentissage des paramètres du modèle

L'apprentissage des paramètres du modèle consiste à définir une méthode d'estimation automatique des paramètres en exploitant les caractéristiques d'une base de données test d'apprentissage. Une solution à ce problème est fournie par l'algorithme de Baum-Welch dans sa version probabiliste [Bau72]. Nous proposons dans ce travail de proposer une version crédibiliste de cet algorithme afin de l'exploiter pour l'apprentissage des paramètres de notre MMC pour la reconnaissance de scénario médical. Pour un modèle λ_{Cr} de MMC crédibiliste quelconque, il s'agit de chercher les paramètres de λ_{Cr} qui maximise la probabilité. La Figure 4.4 est une illustration des différentes phases de l'algorithme de Baum-Welch.

Cependant, il faut noter que la complexité de l'algorithme de Baum-Welch dans sa version crédibiliste que nous présentons dans ces travaux est plus importante que dans sa version probabiliste. Ceci est dû aux fonctions de croyances qui n'ont pas les mêmes propriétés mathématiques que les probabilités et de plus, les croyances sont définies sur les espaces de grande taille. Généralement trois paramètres sont appris pendant la phase d'apprentissage.

- la matrice de transition crédibiliste ;
- la connaissance de l'a priori sur les états à l'instant $t=1$;
- les modèles d'observation liant les observations O_t aux croyances sur les états.

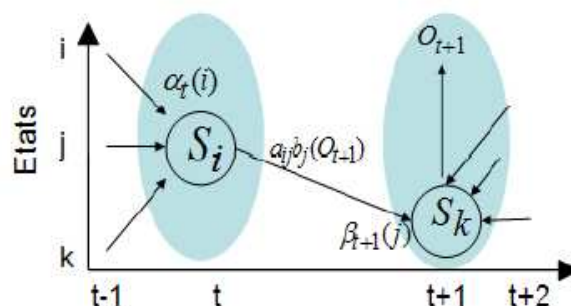


Figure 4.4. Phases de calcul pour la mise en œuvre de l'algorithme de Baum-Welch [Bau72].

La phase d'apprentissage des scénarios dans le cadre de l'estimation et de l'ajustement des paramètres du modèle de Markov est très importante et précède l'étape de reconnaissance proprement dite sans quoi le modèle n'a aucune information sur laquelle un scénario pourrait être reconnu. Deux principales étapes précédées d'une phase d'initialisation permettent de décrire *Baum-Welch crédibiliste* [AME14]:

Initialisation : On définit les paramètres initiaux du modèle en se basant sur les conditions initiales aléatoires. Généralement, on l'initialise à $q_{\alpha}^{\Omega_t=1}(S_i) = 1$.

Etape 1 : Exploitation de la propagation avant (*forward*) qui consiste en un processus de prédiction puis de mise à jour de la prédiction. Cette étape permet d'estimer la variable $q_{\alpha}^{\Omega_t}(s_j)$ de la procédure *forward*.

Etape 2 : Exploitation de la propagation arrière (*backward*), qui permet d'estimer la fonction de croyance a posteriori sur Ω_t par le calcul de masse de croyance avec une règle de combinaison conjonctive. Puis une projection de la masse de croyance combinée (utilisation d'une règle CRC) obtenue sur Ω_{t+1} sur Ω_t et qui nécessite le calcul de $q_{\alpha}^{\Omega_t}(s_j)$ a posteriori. A la fin de cette étape, la variable $q_{\beta}^{\Omega_t}(s_j)$ de la procédure *backward* est obtenue. Une fois les variables $q_{\alpha}^{\Omega_t}(s_j)$ et $q_{\beta}^{\Omega_t}(s_j)$ sont connues, nous pouvons déterminer la variable de lissage crédibiliste $q_{\gamma}^{\Omega_t}(s_j)$ et celle d'apprentissage crédibiliste $q_{\xi}^{\Omega_t}(s_j)$ qui s'expriment par les équations 4.15a et 4.15b suivantes:

$$q_{\gamma}^{\Omega_t} = q_{\alpha}^{\Omega_t} \odot q_{\beta}^{\Omega_t} \quad (4.15a)$$

$$q_{\xi}^{\Omega_t \times \Omega_{t+1}} = q_{\alpha}^{\Omega_t} \uparrow_{\Omega_t \times \Omega_{t+1}} \times q_{\beta}^{\Omega_{t+1}} \uparrow_{\Omega_t \times \Omega_{t+1}} \times q_b^{\Omega_{t+1}} [O_t] \uparrow_{\Omega_t \times \Omega_{t+1}} \times q_{\alpha}^{\Omega_t \times \Omega_{t+1}} \quad (4.15b)$$

Fin: Ré-estimer les paramètres du MMC en utilisant les valeurs mises à jour de A_{Cr} , B_{Cr} et q^{Ω_0} en procédant à une nouvelle itération jusqu'à convergence des paramètres du modèle. *L'algorithme 2* ci-dessous présente les différentes étapes sous forme d'instructions du *Baum-Welch* crédibiliste [AME14].

Algorithme 2 : Procédure de Baum-Welch Crédibiliste $q_{\gamma}^{\Omega_t}(s_j)$

- 1: **Initialisation:** choix des paramètres du MMC
- 2: **Répéter :**
- 3: **Pour** chaque séquence observée:
- 4: **Calculer** $q_{\alpha}^{\Omega_t}(s_j)$ avec la propagation avant (forward) crédibiliste
- 5: **Calculer** $q_{\beta}^{\Omega_t}(s_j)$ avec la propagation arrière (backward) crédibiliste
- 6: **Calculer** $q_{\xi}^{\Omega_t}(s_j)$
- 7: **Calculer** $q_{\gamma}^{\Omega_t}(s_j)$
- Fin Pour**
- 8: Ré-estimer les paramètres du MMC crédibiliste jusqu'à la stabilisation.

4.3.4. Reconnaissance de scénarios crédibilistes

Une fois la phase d'apprentissage des paramètres du modèle MMC terminée, l'étape de reconnaissance proprement dite des scénarios peut ainsi s'effectuer. A cet effet, nous exploitons soit un algorithme de *Viterbi crédibiliste* soit à nouveau les procédures de propagations *forward* et *backward crédibilistes* [RRP07]. Les propagations respectent les mêmes schémas classiques que ceux présentés dans le cadre probabiliste. La Figure 4.5 illustre un modèle graphique représentant les propagations forward et backward crédibilistes entre deux instants où Ω_t est l'ensemble des états.



Figure 4.5. Modèle graphique représentant les propagations forward et backward crédibilistes entre deux instants [Ram07].

4.3.4.1. Propagation *forward* crédibiliste

La propagation *forward crédibiliste* à l'image de celle probabiliste permet de filtrer en partie les fonctions de croyances (les probabilités) en ligne afin de reconnaître une

séquence d'états. En effet, elle s'effectue en trois étapes que sont l'initialisation, l'induction et l'étape de fin.

Initialisation forward crédibiliste : A partir des conditions initiales aléatoires définies, on estime les paramètres initiaux du modèle. Il s'agit donc de forcer en initialisant une variable avec un a priori vide à $t=1$, en utilisant l'équation $\forall S_i \subseteq \Omega_1 \quad q_\alpha^{\Omega_1=1}(S_i)=1$.

Induction forward crédibiliste : A l'étape d'induction $\forall S_j \subseteq \Omega_{t+1}$ avec $1 \leq t \leq T-1$, alors on peut déterminer ainsi la variable $q_\alpha^{\Omega_t}(s_j)$ forward qui est une communalité issue d'une combinaison d'informations en utilisant l'équation 4.16 :

$$q_\alpha^{\Omega_{t+1}}(S_j) = \left[\sum_{S_i \subseteq \Omega_t} m_\alpha^{\Omega_t}(S_i) \times q_\alpha^{\Omega_{t+1}}[S_i](S_j) \right] \times q_b^{\Omega_{t+1}}[O_t](S_j) \quad (4.16)$$

A chaque instant t , l'induction génère un processus de spécialisation qui permet le transfert de la masse sur les ensembles d'états de faible cardinalité et en conséquence renforce la distribution de croyance. Mais à un instant $t+1$, la CRC avec les observations génère un conflit lorsque $S_i \cap S_j = \emptyset$.

Fin : La décision du choix du meilleur modèle se base sur le conflit généré par la combinaison de CRC utilisée lors de la phase de l'induction pour combiner les observations. La séquence optimale est choisie en minimisant le conflit sur l'ensemble de la séquence d'observation en exploitant l'équation 4.17:

$$L_c^1(\lambda_\alpha) = \frac{1}{T} \sum_{t=1}^T \log(m_\alpha^{\Omega_t}[\lambda_\alpha](\Phi)) \quad \text{et} \quad \lambda_\alpha^* = \arg \min_k L_c^1(\lambda_k) \quad (4.17)$$

L'algorithm 3 détaille les principales lignes d'instructions d'une procédure de *forward crédibiliste d'un modèle de Markov caché crédibiliste*:

Algorithm 3 : Procédure forward crédibiliste $q_\alpha^{\Omega_t}(s_j)$

1: **Pour** i allant de 1 à N faire : $q_\alpha^{\Omega_1=1}(S_i)=1$

2: **Fin Pour**

3: **Pour** t allant de 1 à $T-1$ faire :

4: **Pour** i allant de 1 à N faire: $q_{\alpha}^{\Omega_{t+1}}(S_j) = \left[\sum_{S_i \in \Omega_t} m_{\alpha}^{\Omega_t}(S_i) \times q_{\alpha}^{\Omega_{t+1}}[S_i](S_j) \right] \times q_b^{\Omega_{t+1}}[O_t](S_j)$

5: **Fin Pour**

6: **Fin Pour**

7: **Calculer** : $L_c^1(\lambda_{\alpha}) = \frac{1}{T} \sum_{t=1}^T \log(m_{\alpha}^{\Omega_t}[\lambda_{\alpha}](\Phi))$ et $\lambda_{\alpha}^* = \arg \min_k L_c^1(\lambda_k)$

4.3.4.2. Propagation *backward* crédibiliste et temporelle

L'estimation des paramètres du MMC dans le cadre crédibiliste est effective par la détermination de trois variables auxiliaires que sont la variable *backward* β et les variables γ et ξ qui sont obtenues à partir des variables *backward* et *forward*. Comme dans les modèles de Markov probabilistes, on peut utiliser la variable de propagation avant (de probabilité a posteriori d'états donnés d'observations) pour la reconnaissance de l'état e . En effet, un procédé semblable peut être utilisé dans MMC crédibiliste [SRZ11] et, par conséquent, la propagation de retour génère la variable $q_{\beta}^{\Omega}(S_j)$ lors du processus de prédiction et de fusion qui est estimée avec l'équation 4.18 :

$$q_{\beta}^{\Omega_t}(S_i) = \sum_{S \subseteq \Omega_{t+1}} \left((m_{\beta}^{\Omega_{t+1}} \otimes m_b^{\Omega_{t+1}}[O_t])(S_j) \times q_{\alpha}^{\Omega_t}S_j \right) \quad (4.18)$$

Initialisation *backward* crédibiliste : A ce niveau, il faut définir les paramètres du modèle en se basant sur les conditions initiales aléatoires. Par exemple, elle peut être initialisée à $q_{\alpha}^{\Omega_{t=1}}(S_i) = 1$. Les deux principales étapes suivantes de l'algorithme permettent de mettre à jour les paramètres du MMC de façon itérative jusqu'à convergence, en suivant la procédure décrite à partir des deux étapes ci-dessous:

Induction *backward* crédibiliste : D'une part, l'induction permet d'estimer la variable *backward* crédibiliste $q_{\beta}^{\Omega_t}(S_j)$ avec l'équation 4.19 :

$$\forall S_i \subseteq \Omega_t, q_{\beta}^{\Omega_t}(S_i) = \sum_{S \subseteq \Omega_{t+1}} \left((m_{\beta}^{\Omega_{t+1}} \otimes m_b^{\Omega_{t+1}}[O_t])(S_j) \times q_{\alpha}^{\Omega_t}S_j \right) \quad (4.19)$$

D'autre part, au cours de cette étape la variable de lissage crédibiliste (filtrage de fonctions de croyance hors ligne) qui est définie avec une fonction communalité et la

variable d'apprentissage crédibiliste $q_{\xi}^{\Omega_t}(s_j)$ est aussi estimée. La variable crédibiliste de ré-estimation $q_{\gamma}^{\Omega_t}(s_j)$ est obtenue en exploitant l'observation jointe des variables crédibilistes et *forward* et *backward* via l'équation 4.15a obtenue $\forall S_i \subseteq \Omega_t$. De même la variable de lissage $q_{\xi}^{\Omega_t}(s_j)$ est obtenue en utilisant l'équation 4.15b.

L'*algorithme 4* ci-dessous détaille les principales lignes d'instructions d'une procédure *backward* d'un MMC crédibiliste :

Algorithme 4 : Procédure backward Crédibiliste $q_{\beta}^{\Omega_t}(s_j)$

1: Pour i allant de 1 à N faire : $q_{\beta}^{\Omega_{t+1}}(s) = 1, \forall S \subseteq \Omega_{t+1}$

2: **Fin Pour**

3: **Pour** t allant de 1 à $T-1$ faire : $q_{\beta}^{\Omega_t}(S_i) = \sum_{S \subseteq \Omega_{t+1}} ((m_{\beta}^{\Omega_{t+1}} \otimes m_b^{\Omega_{t+1}}[O_t])(S_j) \times q_{\alpha}^{\Omega_t}S_j)$

4: **Pour** i allant de 1 à N faire:

5: **Fin Pour**

6: **Fin Pour**

7: **Calculer** : $L_c^2(\lambda_{\beta}) = \frac{1}{T} \sum_{t=1}^T \log(m_{\beta}^{\Omega_t}[\lambda_{\beta}](\Phi))$ et $\lambda_{\beta}^* = \arg \min_k L_c^2(\lambda_k)$

4.4. Application à la reconnaissance des scénarios d'actes «potentiellement criminel» et «criminel»

Dans cette section, nous nous proposons de tester nos algorithmes de reconnaissance des scénarios de comportement d'humain «potentiellement criminel» et «criminel» sur une base de séquences d'images test proposées par PETS 2014 bien connue de la communauté de vision par ordinateur. Les séquences testées sont collectées à partir d'un système de vidéosurveillance multicaméras. La Figure 4.6 et la Figure 4.7 montre les images du système de vidéo surveillance multicaméras proposé par PETS 2014. Quatre caméras sont exploitées pour surveiller un environnement routier comme illustre à la Figure 4.6 et la Figure 4.7.



Figure 4.6. Illustration du contexte de surveillance multicaméras (4 caméras) proposé par la base de données test PETS 2014. Cas de scénario de comportement «potentiellement criminel».



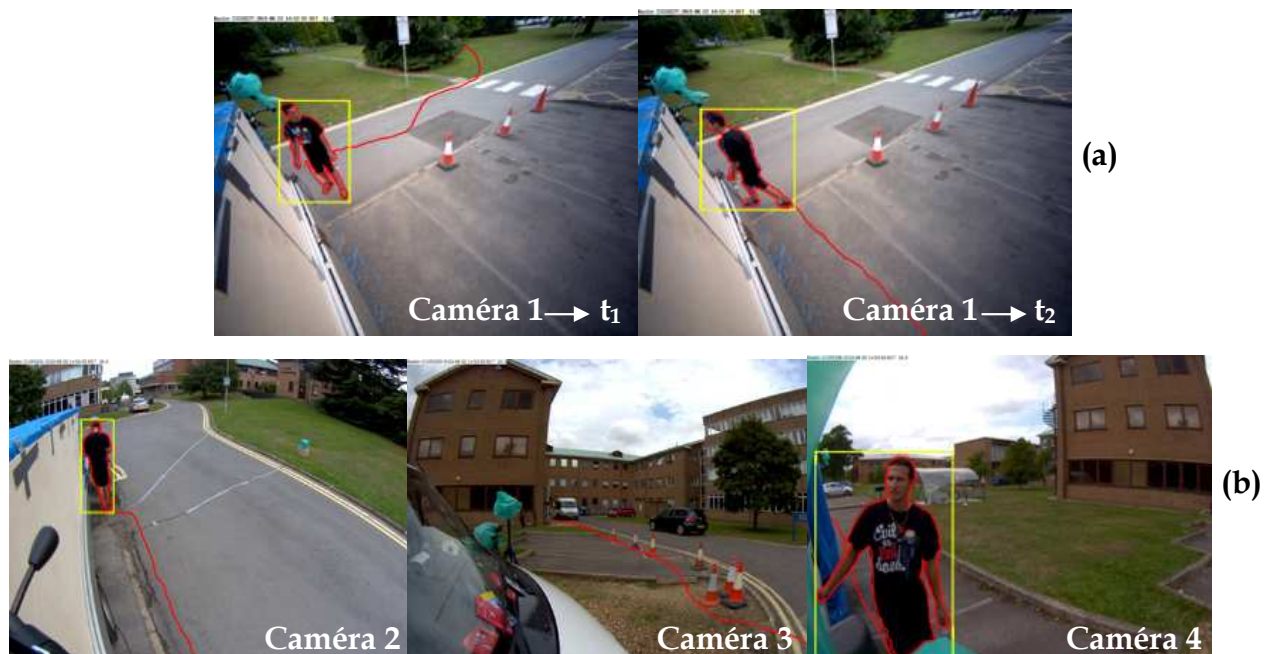
Figure 4.7. Illustration du contexte de surveillance multicaméras (4 caméras) proposé par la base de données test PETS 2014. Cas de scénario de comportement «criminel».

Pour reconnaître chacun des deux scénarios étudiés, un MMCHT de type crédibiliste a été utilisé afin de gérer de façon plus flexible et efficace les imperfections liées aux données. Nous entendons par MMCHTC un modèle de Markov caché crédibiliste. La même technique exploitée au niveau des MMCHT probabilistes est utilisée pour reconnaître les comportements d'humain suspect qui sont de potentiels actes criminels ou tout simplement des actes criminels. La gestion des imperfections concerne la prise en compte de la nature incomplète des données d'apprentissage et la gestion des incertitudes liées tant bien au capteur d'acquisition qu'à la qualité du

résultat des traitements de vision du système de reconnaissance. Pour cela, nous combinons dans une masse de croyance les informations telles que le degré de confiance alloué à la caméra, la probabilité de la durée de l'état du modèle et la probabilité de transition d'état qui permet de former une matrice de transition crédibiliste.

La valeur du paramètre qui représente le degré de confiance alloué à la caméra dépend de la qualité de la détection et du suivi de l'objet via cette caméra. Son estimation est basée sur la connaissance des paramètres de calibration de la caméra, la taille de l'objet et la profondeur.

La durée d'état est calculée en utilisant les distributions de probabilités proposée au chapitre 3 pour modéliser le temps de séjour. A partir de la matrice de transition probabiliste et des deux informations que sont le degré de confiance, la durée nouvelle matrice de transition est ré-estimée avec la probabilité conditionnel afin de construire la matrice de transition crédibiliste. Les procédures d'estimation de différentes est décrite en chapitre 5 et un exemple est fourni en Annexe C.



La Figure 4.8 montre les résultats du système de la reconnaissance de comportement potentiellement criminel d'un homme «un piéton ouvre le véhicule puis le referme» dans la scène. Une analyse globale du système de calibrage multicaméras montre que seuls les

champs de vues de la caméra 1, 2 et de la caméra 4 couvrent correctement la scène mais par contre le champ de vue de la caméra 3 ne couvre pas la scène de l'environnement surveillé (Figure 4.8a et Figure 4.8b). La caméra 1 perçoit un piéton qui avance vers le véhicule et qui arrive très proche du véhicule et y reste pendant un laps de temps de 3s puis il revient au près du véhicule en tournant autour pendant un laps de temps de 4s (images 1012 à 1080). La caméra 2 ne couvre pas toute la scène mais perçoit un homme qui rode autour du véhicule pendant un laps de temps 5s (images 2035 à 2089). La caméra 4 perçoit le piéton avancé tout en s'approchant du véhicule et reste au près du véhicule pendant un laps de temps d'environ de 10s puis tente de forcer pour ouvrir le véhicule (images 1517 à 1643).

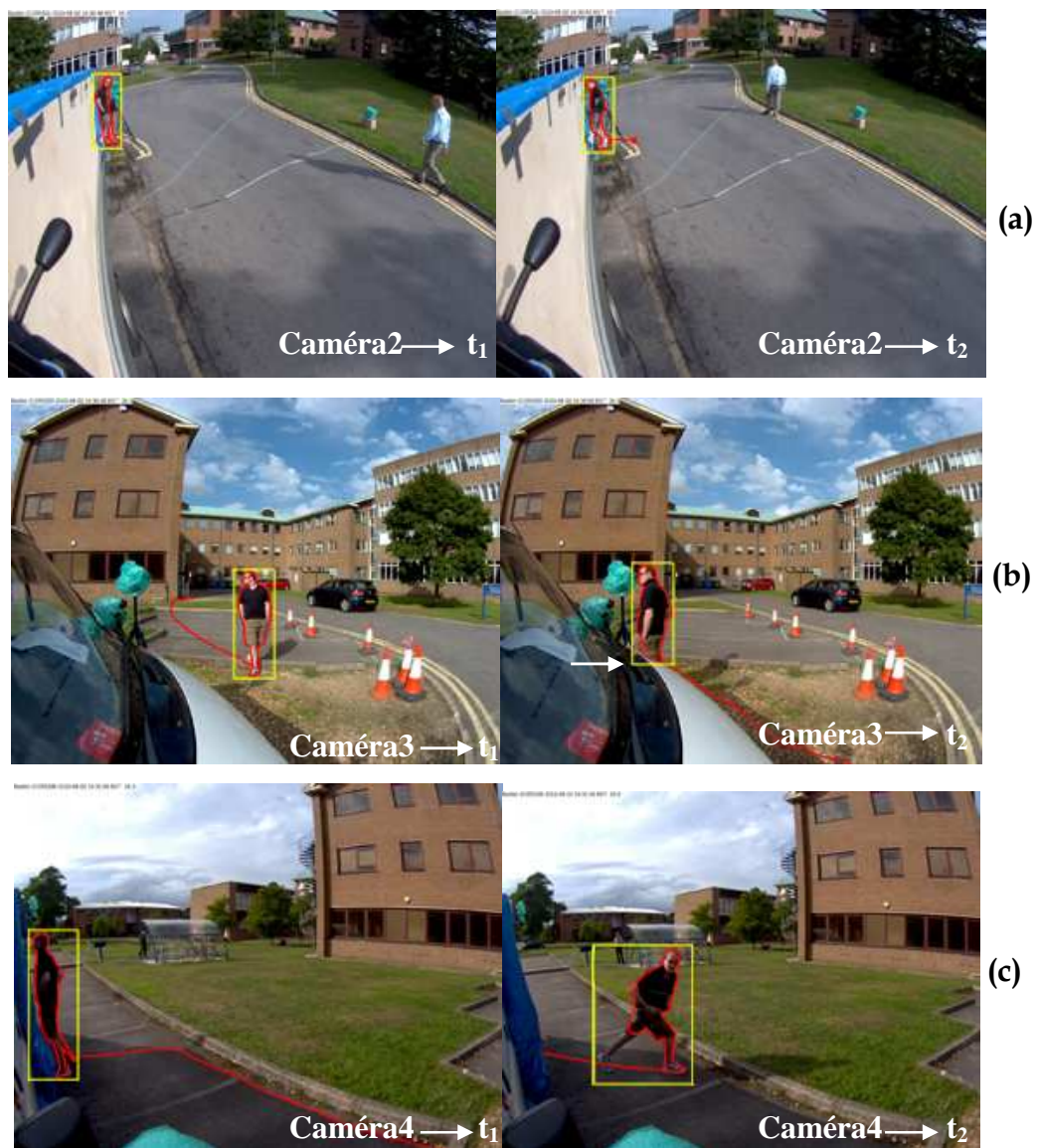


Figure 4.9. Reconnaissance d'événements de comportements «criminel» dans des séquences vidéo tests de la base de données PETS' 14.

La Figure 4.9 montre les résultats du système de reconnaissance pour un comportement criminel d'un homme «un piéton rode autour du véhicule, l'ouvre et vole un ordinateur puis s'enfuit en courant» dans la scène. De même, en analysant le système de calibrage multicaméras on constate que seuls les champs de vues des caméras 2, 3 et 4 couvrent correctement la scène tandis que le champ de vue de la caméra 1 ne couvre pas la scène de l'environnement surveillé (Figure 4.9a, 4.10b, 4.9c). La caméra 2 perçoit un piéton qui avance et arrive très proche du véhicule y reste pendant un laps de temps de 3s puis il revient au près du véhicule en tournant autour pendant un laps de temps de 4s (images 1062 à 1080). La caméra 3 détecte un piéton arrive du bâtiment et qui avance vers le véhicule puis passe très proche devant le véhicule avec un temps moyen de 7s (images 1630 à 1742). Il tourne autour du véhicule pendant temps moyen de 8s (images 1630 à 1742). La caméra 4 montre un piéton tout proche du véhicule et rode autour du véhicule pendant un laps de temps d'environ 10s tout en regardant à gauche et à droite pendant puis ouvre la porte du véhicule pour prendre un objet sous forme rectangulaire et enfin s'enfuit en courant (images 1608 à 1643).

4.5. Apport du MCT pour la gestion d'imperfections

Malgré la puissance, l'efficacité et les bonnes performances enregistrées dans diverses applications par l'exploitation des MMCs probabilistes, quelques limites du modèle subsistent. Ces limites sont essentiellement relatives au cadre de raisonnement qui est le cadre probabiliste. Ce cadre de raisonnement est moins adapté à certaines situations réelles en rapport avec la nature des données et de leur source. Face à ces lacunes identifiées dans le modèle probabiliste, celle crédibiliste constitue une réponse générique à ces problèmes. Par exemple, un MMC probabiliste n'est pas adapté dans le cas où les transformations numérique-symbolique des observations en états ne sont pas toutes probabilistes. De même lorsque les transitions entre états ne sont pas toutes des probabilités conditionnelles le modèle de Markov standard ne répond pas et c'est ainsi que le MCT de par son caractère générique, permet de combiner des informations venant de sources disparates et différentes, etc.

Dans le modèle MMC classique, la connaissance des distributions sur les états initiaux est plus ou moins obligatoire, sinon dans le cas contraire, la théorie des probabilités se propose d'utiliser une hypothèse d'équiprobabilité. Cette dernière, consiste à attribuer la même probabilité à chaque état ce qui n'est pas toujours vérifié et mieux l'ignorance totale est confondue à l'équiprobabilité. Par contre, les fonctions des croyances traitent plus prudemment et mieux cette méconnaissance de la distribution sur les états initiaux en la modélisant par une fonction croyance vide. C'est aussi le cas lorsqu'il y a d'ambigüité sur les transitions entre deux états à un même instant ou entre deux états consécutifs. Cependant, le cadre générique que constitue le MCT permet de prendre en charge l'ambigüité en le modélisant par une fonction croyance certaine soit par exemple une masse traduisant l'ambigüité sur la transition entre deux états consécutifs et vaut 1.

Un avantage conséquent des fonctions des croyances est leur efficacité dans le cadre de la fusion ou de la combinaison d'informations. Le cadre de raisonnement dans le MCT est très adapté pour la combinaison d'informations car il tient compte de la nature et de la qualité des sources à fournir des informations fiables ou bruitées. Plusieurs opérateurs très prudents ont été proposés pour combiner des informations de façon efficace sans introduire un biais sur la donnée et facilite ainsi la prise de décision. Les probabilités échouent lorsqu'on doit combiner des distributions non indépendantes et que la théorie des probabilités fait office d'indépendance comme hypothèse tandis que le MCT propose une merveilleuse solution par l'utilisation d'opérateurs prudents.

De nombreux apports significatifs ont été produits sur le système de reconnaissance par l'exploitation des fonctions de croyance qui restent et demeurent depuis des décennies en pleine expansion avec diverses applications sensibles qui nécessitent une gestion efficace de l'incertitude, de l'imprécision et l'incomplétude. Parmi ces apports, nous pouvons citer une fine description du modèle de transition des états, la modélisation du doute, de l'ambigüité et de l'ignorance totale.

Conclusion

Nous avons décrit dans ce chapitre une approche que nous proposons pour la reconnaissance crédibiliste de scénarios qui exploite les modèles de Markov cachés dans le cadre des modèles de croyances transférables. L'approche proposée est basée sur une idée d'adaptation des mécanismes dans le modèle Markov standard probabiliste. Nous exposons une nouvelle formalisation des algorithmes des MMCs dans le cadre des Modèles de Croyances Transférables. Cette reformulation permet de combiner la puissance qui n'est plus à démontrer des modèles très populaires pour diverses applications de Markov cachés, de la flexibilité et de la généralité des outils tout aussi puissants que robustes du MCT. Cette extension permet de faire face à la modélisation d'un panel de connaissances. En conséquence, ce modèle grâce à ces avantages décrits plus haut fournit un outil générique permettant d'utiliser les connaissances telles qu'elles existent, que ce soit des probabilités, des possibilités ou des fonctions de croyances. L'aspect de l'évolution temporelle des séquences d'états (enchaînement d'événements élémentaires) est tout aussi abordé dans notre approche par une estimation explicite du temps passé par le modèle dans un état sous forme de probabilité qui sera transformée par la suite en une masse de croyance. Ce dernier est combiné pour former une matrice évidentielle de transition des états pour être transformée en communalité. La structure hiérarchique des événements est aussi un aspect qui nous a intéressés dans le modèle proposé.

Chapitre 5

Reconnaissance d'activités dans des vidéos médicales et expérimentations

Introduction

L'ensemble des expérimentations effectuées dans le cadre de la mise en œuvre de la chaîne de traitement du système de reconnaissance d'activités humaines en milieu médical est présenté. Tout d'abord, une description globale du système expérimental et de la base de données médicale avec les différentes catégories de scénarios est faite. Ensuite, nous avons présenté les différentes expérimentations liées aux traitements bas niveau qui concernent notamment d'une part, les prétraitements des images médicales en exploitant une technique basée sur une technique de correction de distorsion stéréoscopique via l'estimation de la matrice fondamentale en exploitant la géométrie épipolaire. Et d'autre part, elles sont relatives à la détection et le suivi des objets mobiles dans la vidéo en exploitant une approche basée sur la soustraction de fond et une différence inter-image avec une image de référence. Enfin, nous décrivons les expérimentations liées aux traitements de haut niveau pour l'analyse et l'interprétation intelligente de la scène médicale en exploitant les modèles de Markov cachés hiérarchico-temporels dans le cadre des modèles de croyances transférables. L'ensemble des résultats obtenus pour la reconnaissance de scénario avec l'approche que nous proposons a été présenté et analysé afin de faire ressortir les apports et les limites probables de notre technique.

5.1. Contexte expérimental

Dans cette section, nous présentons une vue d'ensemble et sommaire du système de vision à perception multicaméras proposée dans nos travaux pour compléter le système de surveillance sonore existant en unités de soins intensifs médicaux. Puis, nous décrivons la base des données tests utilisées dans le cadre de nos expérimentations. Et enfin, quelques apports du système proposé d'un point de vue général pour un système de vidéosurveillance et dans un second temps d'un point de vue de surveillance médicale.

5.1.1. Vue d'ensemble du système expérimental

Le développement et le test des solutions de surveillance des activités (états et comportements) des patients et des activités cliniques médicales nécessitent un cadre expérimental réel ou tout au moins proche du contexte réel médical. Pour répondre à ce besoin important pour la validation et l'analyse de l'impact de ces solutions, une salle de soins intensifs a été aménagée à cet effet dans la section cardiologie du Centre National Hospitalier Universitaire Hubert Koutoukou MAGA (CNHU-HKM) de Cotonou au Bénin. La salle de soins intensifs qui sert de site expérimental pour nos travaux de recherches est divisée en deux pièces que sont : la salle de surveillance où on retrouve les infirmiers et les médecins surveillants et la salle de soins intensifs proprement dite où sont hospitalisés les patients. Cette dernière est séparée de la première par une vitre à travers laquelle se fait une surveillance visuelle des médecins surveillants. Cet état de chose est d'ailleurs l'un des premiers facteurs inspirant la proposition d'un système de surveillance visuelle intelligente en assistance pour ces derniers.

La mise en place de cette salle expérimentale permet aussi d'explorer la construction des scénarios intéressants (événements médicaux importants) à reconnaître puis aide à la découverte de nouveaux scénarios par analyse de la base de vidéos recueillies. D'autre part, elle permet de tester, de mettre en œuvre et d'évaluer la performance des approches proposées. Deux principales catégories de capteurs (capteurs d'informations médicales) sont installées dans cette salle à savoir: les capteurs de signaux physiologiques du patient puis les capteurs de vision pour l'analyse des

comportements du patient. La salle intelligente médicalisée est équipée de six lits de patients, et au niveau de chaque lit, nous pouvons retrouver l'ensemble des équipements de surveillance en USIs de base tel que le moniteur électrocardiographique deux traceurs minimum reliés par câble ou par télé-métrie à une unité centrale intégrant un système d'alarmes, un enregistreur automatique et un système de stockage d'au moins 24 heures. En complément à ce système de surveillance de base, chaque lit est équipé d'une caméra calibrée de façon à ce que son champ de vision couvre la totalité de la surface du lit pour permettre d'avoir une vue globale du patient quelle que soit sa position sur le lit médicalisé (couché ou assis). L'ensemble de la salle est organisé comme présenté à la Figure 5.1. & 5.2.

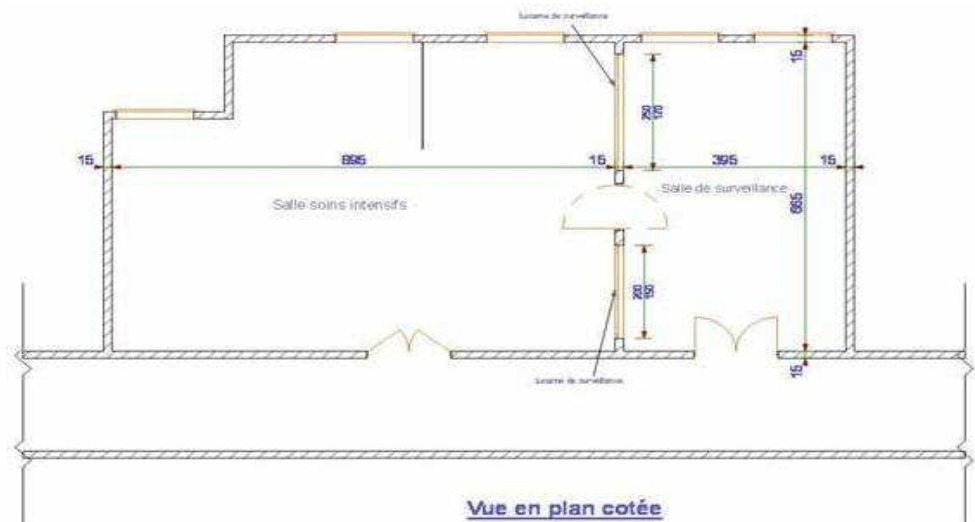


Figure 5.1. Vue en plan côté de la salle de soins intensifs du système expérimental.

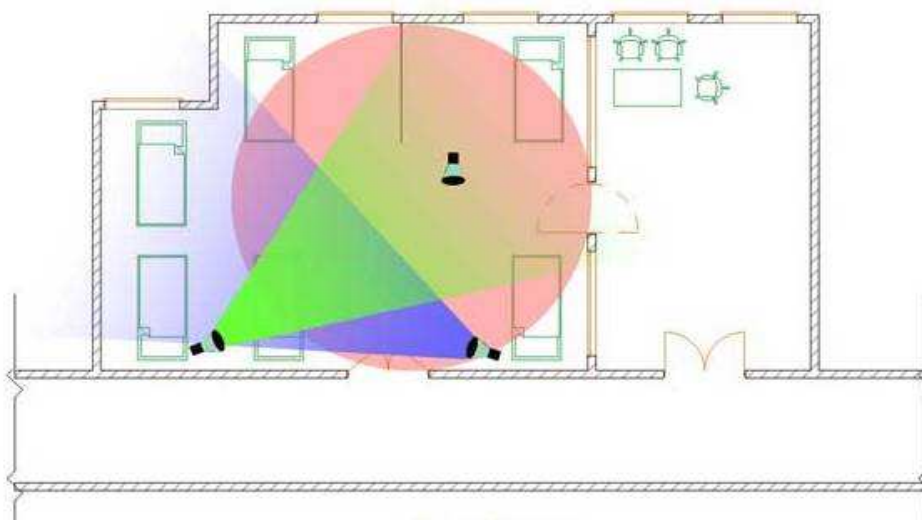


Figure 5.2. Illustration du calibrage de système de vision dans la salle de soins d'USIs.

5.1.2. Description du système visuel expérimental proposé pour le monitoring des patients en USIs

Globalement, les patients sont monitorés dans la salle de soins via un système de vision composé de huit caméras dans notre cadre expérimental et illustré comme à la Figure 5.2. Le système de monitoring visuel est qualifié d'intelligent par sa capacité d'analyse et d'interprétation automatique des activités médicales humaines (praticiens hospitaliers et patients). Il constitue de ce fait, une aide à la surveillance intelligente en temps réel des patients, en amont et en aval une aide à la décision médicale. Nous avons mis en place un système de surveillance visuelle (Figure 5.2) permettant de reconnaître des actions et des activités humaines dans les vidéos contenant des scènes médicales. Une action est interprétée comme un état tandis qu'une activité est représentée par une séquence (un enchaînement) d'états qui représentent les actions de base ou événements élémentaires.

L'entrée du système est une séquence d'images où les objets surveillés sont absents (salle de soins vide) ou tout au moins les objets sont supposés statiques (sans mouvement) à la première image (celle qui sert de référence pour la détection des mouvements d'objets mobiles). Les paramètres du mouvement sont au prime abord estimés afin de détecter les objets mobiles.

Ensuite, un algorithme de catégorisation basé sur la segmentation à base d'attributs morphologiques nous permet de classer les objets mobiles dans la vidéo en deux classes : les patients et les médecins et si un objet n'est pas détecté appartenant à l'une au moins de ces deux classes, il est déclaré appartenant à une troisième qui est celle des assistants des patients ou des visiteurs dont les heures de visite strictement réglementées définies et connues a priori. Ainsi, par classe d'objets mobiles détectés et suivis, des caractéristiques locales et globales sont extraites et obtenues pendant la phase d'extraction de caractéristiques puis sont transformées en croyance sur les actions (action dans une région d'intérêt prédéfinie) qui vont être enfin combinées dans le cadre du MCT afin de gérer l'incertitude, l'imprécision générée sur les données lors des traitements de vision bas niveau tout en tenant compte de l'incomplétude des connaissances médicales.

Enfin, la sortie du système est une décision de reconnaissance des scénarios prédéfinis par une connaissance d'expert médical. Cette décision est basée sur le

Modèle de Markov Caché Crédibiliste (MMCCr), qui permet de reconnaître les scénarios. La Figure 5.3 présente l'architecture d'un point de vue global de toute la chaîne du système de reconnaissance.

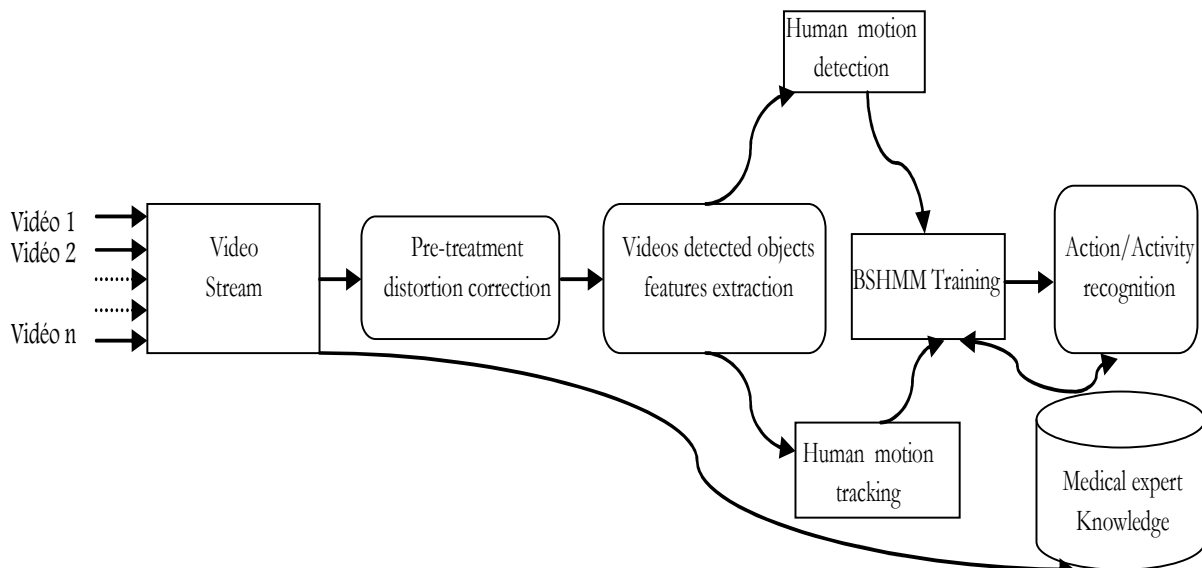


Figure 5.3. Architecture globale de chaîne de traitement du système de reconnaissance [AME13].

5.1.3. Description de la base de vidéos médicales d'USIs exploitées

D'importants tests et expérimentations ont été réalisés sur les données de vidéos médicales recueillies à partir du système de surveillance visuelle en production expérimentale en salle de soins intensifs au service cardiologique du CNHU-HKM.

5.1.3.1. Caractérisation de la base de vidéos médicales

La scène vidéo est capturée par différentes caméras auto-calibrées et installées dans la salle de soins. Un module de suivi multicaméras est utilisé pour détecter les mouvements afin de fournir la liste des positions des personnes (le patient et le médecin) seules en coordonnées (x,y) puis des caractéristiques locales des régions d'intérêts prédéfinies. Nous avons collecté et utilisé pour nos expérimentations 47 séquences vidéo constituées des séquences de longueur de 3000 images. Chaque séquence d'observation est enregistrée toutes les secondes et demi de la vidéo à 20fps. La durée des séquences vidéo est de 150 secondes. La base de données vidéo se caractérise par son hétérogénéité avec une forte variation de l'éclairage (jour et nuit) et

la présence dans la scène de plusieurs personnes comme les médecins et les patients (hommes, femmes avec des actions différentes, des couleurs de peau variées). On peut aussi noter la présence d'autres personnes qui se déplacent telles que les visiteurs, entre autres qui peuvent apparaître à des moments bien définis et connus. Quelques images de la base de données sont fournies par catégories d'objets à la Figure 5.4.

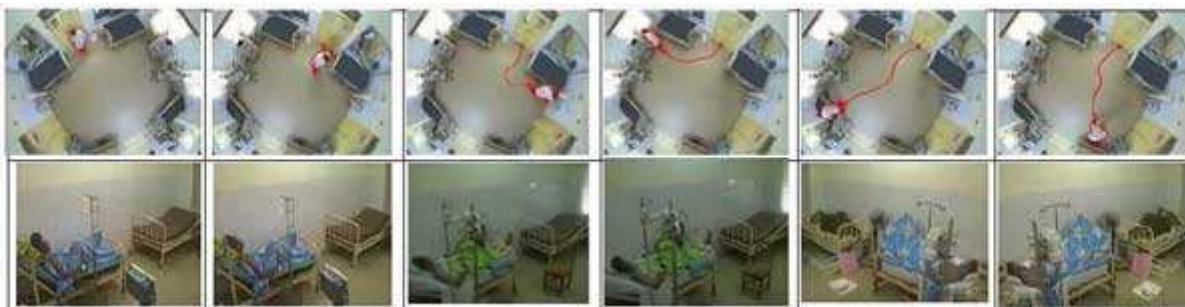


Figure 5.4. Exemples d'images intégrant des scénarios reconnus par notre système.

La première colonne de séquences d'images représente la séquence illustrant un scénario lié au médecin qu'est «**la visite médicale d'un patient**» puis la séquence de la seconde colonne représente un scénario lié au patient qu'est «**l'agitation du patient**».



Figure 5.5. Illustration d'exemples de séquences d'images collectées par caméra et par catégorie d'objet.

5.1.3.2. Description et modélisation des scénarios à reconnaître

Deux classes d'activités nous ont intéressé et sont définies par rapport aux deux principaux objets (le patient et le médecin) dont les comportements sont analysés dans la scène à savoir :

- **Activités du patient :** agitation du patient, l'émission de douleur par le patient et convulsion du patient.
- **Activités du médecin :** visite médicale d'un patient, tour de contrôle d'état clinique des patients, et soins médicaux à un patient.

Pour chaque classe d'activités, les trois types de scénarios sont testés et reconnus par notre système de reconnaissance de scénario. S'agissant de la classe «*les activités du patient*», nous avons testé les scénarios suivants : scénario 1 (*agitation du patient*), scénario 2 (*émission de douleur par le patient*), scénario 3 (*convulsion du patient*).

- **Le modèle de scénario 1:** *agitation du patient*, est décrite et modélisée par trois actions de base qui sont des événements élémentaires : mouvement (léger et/ou saccadé) du pied (gauche et/ou droit) du patient et/ou le mouvement (léger et/ou saccadé) de la main (gauche et/ou droite) du patient. En fonction de la succession et de la cadence, en d'autre terme le rythme du mouvement des membres du patient et ses mouvements globaux corporels, nous décidons qu'un scénario «*le patient s'agite*» est détecté. A cet effet, mouvement des mains respectivement gauche et droite est symbolisé « H_L » appelé «*Left Hand motion*» et « H_R » est à son tour même nommé «*Right Hand motion*». De même, s'agissant du mouvement du pied gauche respectivement du pied droit, il est symbolisé « F_L » appelé «*Left Foot motion*» et « F_R » nommé «*Right Foot motion*».
- **Le modèle de scénario 2:** *l'émission de douleur par le patient*, est modélisé non seulement sur la base des mouvements de membres (les mains et les pieds) du patient mais aussi sur les mouvements de sa tête et même de façon plus efficace en combinant ses expressions faciales (émotions, rigoles, renfroge la mine). Ce dernier est assez complexe à modéliser et nécessite aussi un capteur d'acquisition vidéo sophistiqué très proche du visage du patient. Nous avons combiné les mouvements des membres à ceux de la tête pour décider d'une détection d'émission de douleur par le patient tout en faisant une analyse de texture avec l'histogramme du visage avec filtre détecteur de visage. En complément à la modélisation du mouvement des membres du patient dans la section précédente, deux états sont utilisés pour reconnaître les mouvements de tête (tête à gauche : H_L et tête à droite : H_R). La combinaison de la reconnaissance des deux mouvements permet de décider s'il y a une émission de douleur par le patient.

- **Le modèle de scénario 3:** *convulsion du patient*, pour ce scénario un système de reconnaissance assez grossière est utilisé pour sa reconnaissance. Nous exploitons simplement le modèle de reconnaissance du scénario «*agitation du patient*» en y intégrant une contrainte temporelle afin de reconnaître ce scénario. En effet, selon les connaissances expertes du domaine médical, la convulsion se manifeste par plusieurs signes physiologiques chez le patient généralement visible dans les comportements et actions du patient. Pour une détection grossière de ce scénario, nous avons exploité un modèle simple basé sur une approche d'analyse des mouvements et attitudes du patient en fonction du rythme et de la cadence des mouvements du patient. A cette première phase d'interprétation doit s'ajouter une seconde analyse un peu plus raffinée des comportements du patient afin de détecter de façon plus efficiente la convulsion. En d'autres termes, la reconnaissance efficace de ce scénario nécessite une modélisation plus complexe et aussi des capteurs d'acquisitions d'images plus sophistiquées et proche du patient. Mais dans nos travaux actuels, nous avons effectué une analyse haut niveau et globale des comportements du patient en choisissant de modéliser ce scénario sur la base de l'enchaînement des mouvements des membres (les mains et les pieds) du patient mais aussi tout son corps tout en mettant en exergue le rythme, la cadence et la vitesse des mouvements corporels de ce dernier.

S'agissant de la classe «*activité du médecin*» nous avons les scénarios suivants : scénario 4 (*visite médicale du patient*), scénario 5 (*tour de contrôle d'état clinique des patients*), scénario6 (*soins médicaux à un patient*).

- **Le modèle de scénario 4 :** *visite médicale du patient*, est divisé en six sous-scénarios qui sont regroupés dans la présence du médecin aux différents lits des malades : «*présence du médecin au Lit1*», «*présence du médecin au lit 2*», jusqu'au sous-scénario «*présence du médecin au lit N*». Par exemple le scénario «*présence du médecin au lit1*» est construit à partir des caractéristiques de l'objet médecin détecté mobile, de son suivi et des régions d'intérêts définies à base d'attributs de niveau de gris et de texture auquel est associé symboliquement des alphabets (A, B et C) intégrés dans un codebook. L'élaboration des événements élémentaires est basée sur la détection de la présence du médecin et ses actions primitives dans la

région d'intérêt prédéfinie et dont l'enchaînement correspond au scénario important à reconnaître.

- **Le modèle de scénario 5:** *tour de contrôle d'état clinique des patients*, est qualitativement du même type que le scénario1, mais diffère de ce dernier par le temps passé par l'objet médecin dans un état donné. D'autre part, le tour du médecin exige la présence du médecin dans six régions spécifiques que sont celles définies autour du lit du patient et peut commencer par n'importe quel lit de patient.
- **Le modèle de scénario 6:** *Les soins médicaux à un patient*, est divisé en six sous-scénarios identifiés par la présence du médecin devant un lit spécifique du patient : «soins médicaux au patient du Lit1», «soins médicaux au patient du Lit2», jusqu'au scénario «soins médicaux au patient du Lit6». Par exemple le scénario «présence du médecin au lit 1» est caractérisé par les actions du médecin en face du patient avec une durée spécifiquement prédéfinie pour ces soins. En effet, le temps passé par le médecin chez un patient pour faire un soin est un a priori car il est bien défini, connu cliniquement dans les conditions normales et varie en fonction du type de soin. Le Tableau 5.1 résume les différents types de scénarios reconnus en se basant sur les classes d'objets présents dans la scène.

Noms des scénarios	Scénarios Ordinaires (SO)	Scenarios Complexes (SC)	Définition des types de scénario par classes d'objets de la scène
Le patient s'agite	---(c)/(d)		Scenario 1 : Action du Patient (AP)
Le patient ressent des douleurs		-- H _L -H _R - --- (c)	Scenario 2 : Action du Patient (AP)
Le patient convulse		-H _L -H _R -F _L -F _R ---(c) & (d)	Scenario 3 : Action du Patient (AP)
Le médecin fait des soins à patient d'un lit N	A-C-B --- (a)		Scenario 4 : Action du Médecin (AD)
Le médecin fait une visite à un patient d'un lit N	A-C-D --- (b)		Scenario 5 : Action du Médecin (AD)
Le médecin ou le staff d'infirmier fait un tour de visite de vérification d'état clinique des patients en USIs		A-C-D-C-B-E-F-H-F-G-I-- (a)/(b)	Scenario 6 : Action du Médecin (AD)

Tableau 5.1. Description des différentes classes de scénarios reconnues.

5.2. Prétraitement des images médicales d'USIs

Dans nos expérimentations, les prétraitements concernent substantiellement la correction de la distorsion introduite sur les images collectées à partir des caméras installées au plafond de la salle de soins afin de couvrir globalement toute la salle.

5.2.1. Modèle de prise en compte des distorsions

Les optiques réelles ne se comportent pas exactement comme le modèle de base de la caméra nommée modèle affine ou sténopé, mais elles présentent des aberrations. Le modèle de base de la caméra décrit dans le chapitre 2 peut être amélioré en tenant compte des distorsions non linéaires. Trois différents types de distorsions peuvent être considérés en fonction de la source qui les a produites [Bou00, LZ99a] :

- la distorsion radiale est le plus souvent due à des défauts dans la courbure radiale des lentilles optiques constituant la caméra.
- la distorsion de décentrage due à un mauvais alignement des centres optiques des lentilles de la caméra.
- la distorsion prismatique due à une inclinaison des lentilles les unes par rapport aux autres.

D'un point de vue mathématique, la prise en compte des distorsions se traduit par l'addition d'un terme correctif aux coordonnées caméras des points images issus du modèle projectif $\{x,y\}$ pour obtenir des coordonnées caméras corrigées $\{x_d,y_d\}$.

$$\begin{cases} x_d = x + \delta_x(x, y) \\ y_d = y + \delta_y(x, y) \end{cases} \quad (5.1)$$

Les termes $\delta_x(x, y)$ et $\delta_y(x, y)$ sont eux mêmes la somme de plusieurs termes correctifs correspondants aux différents types de distorsions considérés et au degré d'approximation du modèle choisi.

5.2.2. Approche de correction multi-vue

Nous avons proposé et utilisé une approche de correction de distorsion stéréoscopique basée sur l'estimation de la matrice fondamentale en exploitant la géométrie épipolaire. Pour cela, à partir de points d'intérêts (pixels appartenant à des régions moins distordues) extraits avec un descripteur local dans les images issues de différentes vues chevauchantes du système multicaméras, une correspondance des points dans les images (vue de droite et de gauche) est réalisée.

Nous avons fait recours à la géométrie épipolaire pour estimer la matrice fondamentale. Une fois, la matrice fondamentale déterminée, il est enfin possible d'utiliser les contraintes épipolaires et la relation mathématique qui existe entre la matrice fondamentale et la matrice essentielle pour calculer cette dernière au besoin. A l'issue de tous ces traitements, les paramètres intrinsèques et extrinsèques de la caméra sont déterminés et représentent les variables d'entrées du modèle de correction de distorsion. Une rectification projective, de préférence épipolaire peut être nécessaire si les caméras ne sont pas alignées [ASA⁺00]. L'idée est de rectifier la matrice fondamentale après l'avoir estimé dans le but d'avoir une matrice fondamentale rectifiée pour prendre en compte les déformations induites par le non alignement des caméras. Toutefois, il faut noter que dans notre contexte les caméras sont envisagées alignées, néanmoins nous présentons aussi le cas non aligné.

5.2.2.1. Principe de correction de distorsion multi-vue

Une fois la matrice fondamentale, à partir des outils de la géométrie épipolaire, est estimée et rectifiée au besoin, on peut générer les paramètres du modèle de distorsion qui permettront de corriger les images. Cette distorsion est due à l'optique des caméras, elle ne peut être corrigée que par un traitement de l'image à postériori. Cette phase est indispensable pour pouvoir travailler sur des images le plus proche possible du réel. De plus la distorsion engendrée par l'optique des caméras peut être différente entre la caméra gauche et la droite, il devient nécessaire dans ce cas de corriger cette distorsion pour réaliser un appariement correct par exemple [GD00]. Les fonctions implémentées suivent le schéma synoptique présenté à la figure suivante :

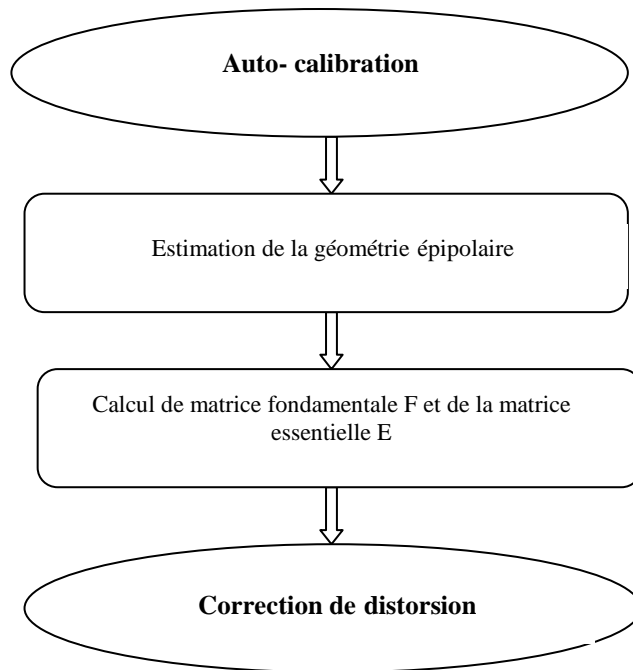


Figure 5.6. Organigramme de correction de distorsion multi-vue [AME⁺12].

L'approche de correction de distorsion proposée réalise de façon élégante et robuste avec des résultats comparables aux approches existantes. Après l'estimation, puis rectification de la matrice fondamentale basée sur la connaissance de la géométrie épipolaire de la scène, les paramètres intrinsèques de la caméra et même extrinsèques sont estimés et au besoin la matrice essentielle. Une fois les paramètres de la caméra connue, les relations mathématiques suivantes permettent de déterminer les coordonnées corrigées, c'est-à-dire non distordues de l'image corrigée. Pour un point de coordonnées $\{x,y\}$ dont les coordonnées caméras corrigées sont $\{x_d,y_d\}$, selon le type de distorsion, nous avons :

Pour une distorsion radiale :

$$\begin{cases} x_d = x + x * (k_1 * r^2 + k_2 * r^4 + k_3 * r^6 + \dots) \\ y_d = y + y * (k_1 * r^2 + k_2 * r^4 + k_3 * r^6 + \dots) \end{cases} \quad \text{avec} \quad r = \sqrt{x^2 + y^2} \quad (5.2)$$

Pour une distorsion décentrage :

$$\begin{cases} x_d = x + [p_1 * (r^2 + 2 * x^2) + 2 * p_2 * x * y] * (1 + p_3 * r^2 + \dots) \\ y_d = y + [2 * p_1 * x * y + p_2 * (r^2 + 2 * y^2)] * (1 + p_3 * r^2 + \dots) \end{cases} \quad \text{avec} \quad r = \sqrt{x^2 + y^2} \quad (5.3)$$

Pour une distorsion prismatique :

$$\begin{cases} x_d = x + s_1 * r^2 + s_1' * r^4 + s_1'' * r^6 + \dots \\ y_d = y + s_2 * r^2 + s_2' * r^4 + s_2'' * r^6 + \dots \end{cases} \quad r = \sqrt{x^2 + y^2} \quad (5.4)$$

Pour une distorsion totale :

$$\begin{cases} x_d = x + k_1 * x * (x^2 + y^2) + p_1 * (3 * x^2 + y^2) + 2 * p_2 * x * y + s_1 * (x^2 + y^2) \\ y_d = y + k_1 * y * (x^2 + y^2) + 2 * p_1 * x * y + p_2 * (x^2 + 3 * y^2) + s_2 * (x^2 + y^2) \end{cases} \quad (5.5)$$

k_1, p_1, s_1, s_2 sont des paramètres qui caractérisent chaque modèle de distorsion.

Il est possible de reconstruire la géométrie épipolaire en estimant la matrice fondamentale à partir de la correspondance entre quelques points et sans information sur les paramètres intrinsèques ou extrinsèques de la caméra utilisée [Fus00]. Chaque correspondance conduit à une équation (5.6) homogène :

$$u \times r' \times F \times u \times l = 0 \quad (5.6)$$

Tous les coefficients de la matrice F (à un facteur près) sont obtenus en établissant au moins huit correspondances ou plus en fonction du niveau de précision recherché et on construit ainsi le système linéaire : $Ax = 0$. La matrice A est de rang 8 et représente une solution unique à un facteur près, c'est à dire proportionnelle à la dernière colonne de V avec $A = UDV'$.

5.2.2.2. Algorithme de correction de distorsion multi-vue

L'algorithme 1, décrit ci-dessous, permet d'estimer la matrice fondamentale F à partir des points témoins choisis et en se basant sur la matrice A construite à travers la résolution du système linéaire ($A = UDV'$) issue de la mise en correspondance des points témoins choisis par exemple par détection de «coins» par la décomposition en valeur singulière.

Algorithme 1 : Estimation de la matrice fondamentale F

- 1: **choisir** la taille du nombre de points à mettre en correspondance ($n \geq 8$)
- 2: **construire** le système: $Ax = 0$ où A est une matrice $n \times 9$ avec $A = UDV'$
- 3: **renforcement** de la contrainte de singularité $Rank(F) = 2$

4: **calculer** la décomposition en valeur singulière de F : $F = \mathbf{UDV}'$

5: **affecter** 0 à la valeur singulière la plus petite soit \mathbf{Dc} la matrice \mathbf{D} corrigée

6: **calculer** à $\mathbf{Fc} = \mathbf{UDcV}'$ qui vaut la matrice \mathbf{F} corrigée (\mathbf{C} 'est la matrice singulière la plus proche de \mathbf{F} au sens de la norme de Frobenius)

L'algorithme de la rectification projective permet de calculer les matrices de rectification afin que les épipôles soient rejetées à l'infini. Ce qui revient à faire un changement de base projectif. Une fois cette rectification effectuée, il est possible de calculer une carte de disparités et une carte des profondeurs. *L'algorithme 2* permet de réaliser la rectification épipolaire tandis que *l'algorithme 3* permet d'estimer la matrice fondamentale \mathbf{F} à base de la géométrie épipolaire.

Algorithme 2 : Rectification projective

1: **choisir** une transformation projective \mathbf{H}' qui projette l'épipôle \mathbf{e}_2 à l'infini

2: **calculer** la matrice \mathbf{H} qui minimise aux sens des moindres carrés la somme des distances $\sum d(\mathbf{H}p, \mathbf{H}'p')$, (avec p et p' les coordonnées de points homologues).

3: **rectifier** les images selon les transformations \mathbf{H} et \mathbf{H}'

Algorithme 3 : Estimation de la matrice fondamentale \mathbf{F}

1: **initialisation:**

Pour chaque pixel de coordonnée (x,y) **Faire** :

$$r_d = \sqrt{x^2 + y^2}$$

Fin pour

2: **estimation de $r_d(x_d, y_d)$:**

Pour i allant de 1 à 10 **Faire** :

$$x_d = \frac{x}{1 + K_1 * r_d^2} \quad \text{et} \quad y_d = \frac{y}{1 + K_1 * r_d^2}$$

Fin pour

3: **transformation inverse :**

Pour chaque pixel de coordonnée (x,y) **Faire** :

$$I(x, y) = I_d(x_d, y_d)$$

Fin pour

5.2.2.3. Expérimentations et résultats

Afin de valider l'approche exploitée, nous avons utilisé plusieurs jeux d'images tests appartenant tant bien qu'à la base de données classiques mise au point par la communauté de vision par ordinateur pour l'évaluation des algorithmes de computer vision et de traitement d'images en stéréoscopie qu'à nos propres bases de données d'images collectionnées du système de vidéosurveillance installé à Unités de Soins Intensifs(USIs) de la cardiologie de l'hôpital. Tout d'abord la méthode a été évaluée à partir d'images de la communauté vision pour comparer nos résultats à ceux des méthodes existantes puis également évaluées sur une paire d'images stéréoscopiques réelles.

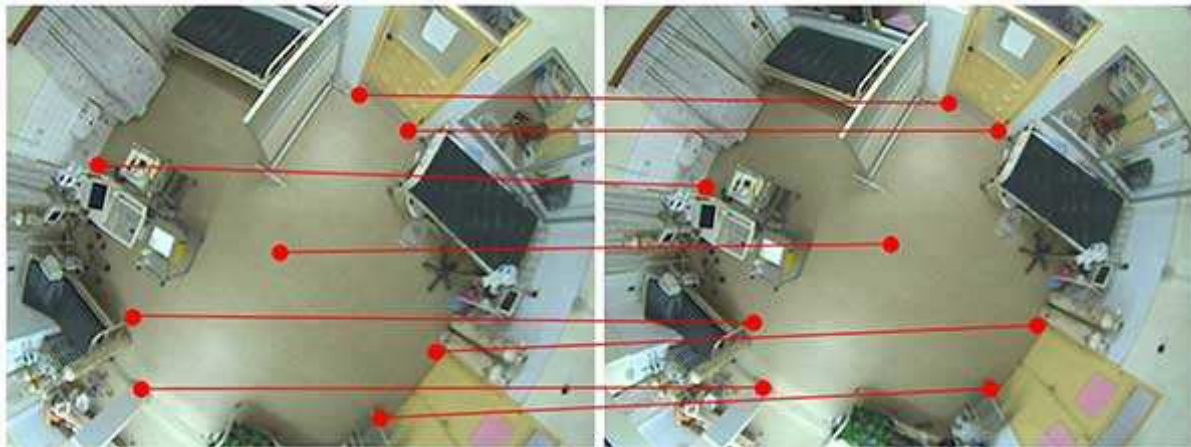


Figure 5.7. Illustration d'exemple d'une paire d'images stéréoscopiques.

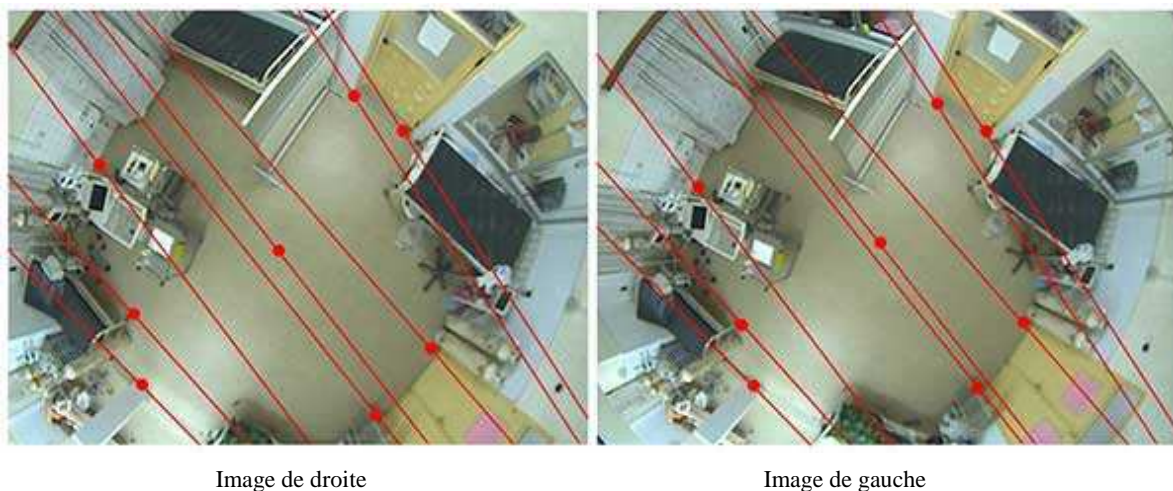


Figure 5.8. Tracé des lignes épipolaires d'une paire d'images stéréoscopiques

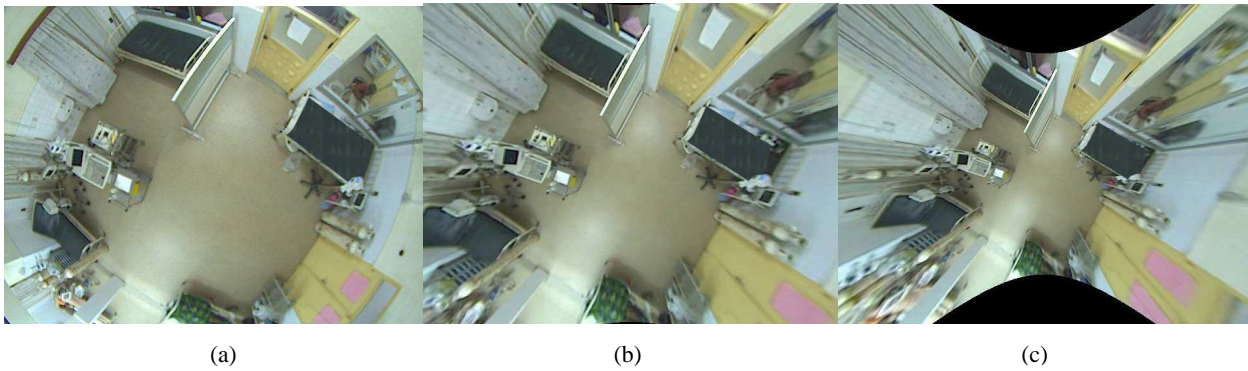


Figure 5.9. Illustration d'exemples d'images distordues corrigées.

La Figure (a) représente l'image distordue avant correction de la distorsion puis la (b) représente une image distordue corrigée à une échelle de 0.7 et enfin (c) une image distordue corrigée à une échelle à 1.

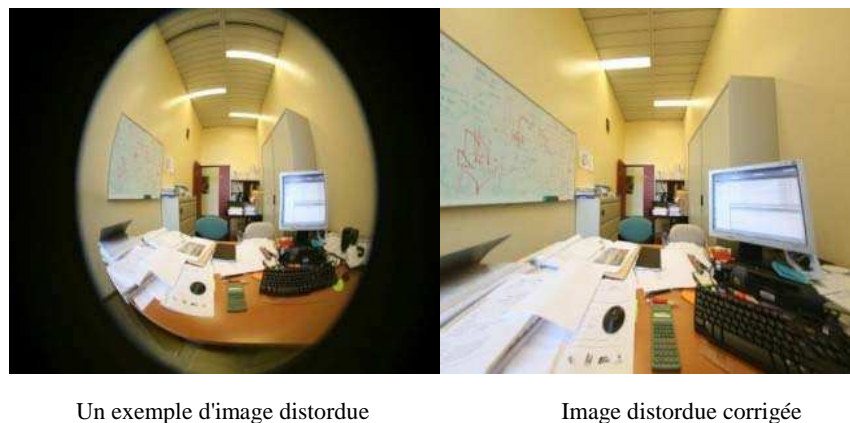


Figure 5.10. Illustration d'exemple d'images distordues corrigées.

La qualité des résultats d'un algorithme en vision par ordinateur que ce soit les traitements de bas niveau ou de haut niveau dépend non seulement de l'efficacité et de la robustesse de l'algorithme mais aussi de l'état ou la qualité des données (bruitées ou non) sur lesquelles l'algorithme est appliqué. La correction des distorsions et la rectification des données images (bruitées) avant l'application de tout algorithme de traitement d'images est obligatoire. Nous avons présenté dans cette section une approche de correction de distorsion avec rectification épipolaire par auto-calibration [HIZ03, NCI02] stéréoscopique à partir d'un système de vision multi caméra. Cette méthode passe par l'estimation de la géométrie épipolaire du système de vision multi

caméras dont les paramètres sont utilisés pour le modèle de correction de distorsion radiale.

Les expérimentations ont montré que cette méthode est particulièrement robuste et élégante du fait de l'auto-calibration et de la technique de détermination des paramètres de la caméra qui permettent d'avoir des valeurs assez précises en entrée pour le modèle de distorsion. De plus, elle se réalise en une seule étape et s'initialise de manière simple. Les résultats obtenus sont très satisfaisants, car en comparaison aux autres méthodes existantes, nous arrivons à être plus précis et à générer moins de déformations projectives.

Cependant, l'initialisation de notre algorithme par rapport aux choix des points témoins reste assez manuelle et il n'existe pas de technique exacte pour les choix de ces points. Dans la littérature on peut noter l'utilisation d'une technique de détection de coins. On peut aussi noter le coût de calcul élevé de notre algorithme bien qu'un compromis ait été trouvé au niveau de la complexité algorithmique de l'utilisation des algorithmes récursifs.

Enfin, une exploitation de la géométrie tri-focale est une voie pour répondre à certaines limites de notre approche.

5.3. Détection et suivi des objets dans les scènes médicales

Nous exposons dans cette section les principaux algorithmes exploités avec les résultats obtenus au niveau de l'étape de détection et de suivi des objets mobiles dans les vidéos de scènes médicales. La plupart de ces algorithmes sont issus des travaux de recherche de C. Motamed [Mot07] dans le cadre de la problématique détection d'objets mobiles en mouvement et de leur suivi dans une séquence vidéo basée sur une image de référence avec une caméra fixe. Notre choix a été guidé par la simplicité et la complexité algorithmique afin de répondre à des aspects temps réel et la possibilité de manipuler des indicateurs crédibilistes permettant d'évaluer la qualité des résultats obtenus à chaque étage du processus de vision. Cette évaluation passe par la définition des indicateurs crédibilistes qui peuvent être fusionnés au besoin dans des fonctions de croyance pour la gestion des incertitudes et des imprécisions générées par ces

traitements bas niveau. Enfin, la valeur de ces indicateurs est propagée à travers les traitements «*haut niveau*» de l'étape d'interprétation de la scène.

L'objectif étant de prendre en compte les imprécisions et les incertitudes afin d'interpréter la scène avec un degré de confiance améliorant de ce fait la qualité de la reconnaissance et permettant d'estimer un taux d'imprécision pour représenter la qualité du système de reconnaissance. Les performances d'un tel système dépendent généralement du contexte et peuvent être comparés à une application particulière. Elles sont relatives à :

- la qualité des décisions,
- compromis entre la qualité et le temps de calcul,
- la robustesse par rapport aux petites perturbations.

5.3.1. Détection des objets en mouvement de la scène médicale avec des indicateurs crédibilistes

L'algorithme de détection exploité est basé sur le principe de la détection de mouvement à partir d'une image de référence qui consiste en une modélisation fixe du fond intégrant une réactualisation contrôlée de la référence avec un seuillage adaptatif pour l'étape de la décision. Les objets en mouvement représentant l'image d'avant plan sont extraits par une simple différence de l'image courante à partir de l'arrière plan [Mot07].

La détection de mouvement, comme toutes les étapes de traitements bas niveau affecte directement le niveau supérieur de l'interprétation. Il devient alors très sensible et important de gérer les conditions affectant la qualité des résultats de cette étape. Par exemple, la détection devient particulièrement sensible lorsque le système doit fonctionner avec des conditions d'éclairage ambiantes. Par conséquent, il importe de tenir compte des conditions affectant l'efficacité et la qualité de notre algorithme de détection de mouvement. La robustesse de l'algorithme réside dans sa résistance face aux problèmes de bases connus du processus de détection de mouvement dont nous avons fait mention au chapitre 2.

Dans notre contexte applicatif, nous avons exploité une approche de détection directe en utilisant une image de référence, car elle est bien adaptée à cause de sa

simplicité et possède une complexité algorithmique intéressante qui permet d'obtenir assez rapidement les résultats à fournir aux étapes suivantes pour l'interprétation automatique de la vidéo. En effet, elle permet de mettre en valeur directement les masques des objets, même s'ils arrêtent leurs mouvements. En contrepartie, par rapport à la différence entre images consécutives, l'obtention et la mise à jour de cette image de référence demeurent une tâche délicate. Le processus de détection du mouvement à partir d'une référence peut être divisé en trois phases :

- la première concerne la génération de la référence;
- la deuxième consiste à estimer une distance entre l'image courante I_k et l'image de référence R_k .
- la troisième concerne la décision en utilisant un seuil de détection qui est adaptatif. Le résultat est alors un masque binaire représentant la projection des objets mobiles sur le plan image.

La génération de la référence. L'image de référence R_k représente une image ou un modèle de la scène exempté d'objets mobiles. Elle peut être construite par la mise à jour d'un modèle à chaque acquisition. Le vieillissement de l'image de référence entraîne de fausses détections et impose une réactualisation adaptée. Idéalement, l'image de référence constitue alors une image de la scène la plus récente ne contenant aucun objet mobile. Cette référence peut être initialisée lorsque cela est possible, de manière simple avec une scène ne contenant aucun objet mobile. Ensuite un filtre de réactualisation du premier ordre est généralement utilisé pour mettre à jour la valeur de chaque pixel P de l'image de référence (5.7). Le paramètre « a » intervient dans la rapidité de la réactualisation.

$$R_k(P) = a \times R_k(P) + (1 - a)I_k(P) \quad (5.7)$$

Estimation de la distance entre l'image courante et celle de la référence. Un opérateur de combinaison est construit pour calculer la distance entre l'observation et la référence de fond.

La décision de détection pour la génération de l'image de premier plan. Il est obtenu par seuillage de la distance entre l'observation et la référence. On note D_k

l'image mise en évidence au premier plan pour les régions de pixels qui ont bougé dans la séquence d'image k . Le seuil de détection η est réglé automatiquement ou défini par l'utilisateur. Il est généralement réglé plus élevé que l'amplitude de bruit typique. Il est aussi possible de régler le seuil le plus complexe en utilisant la connectivité des régions détectées.

Une bonne détection d'objet mobile est une condition nécessaire pour la réussite des autres étapes du processus du système de vision. Ainsi, nous proposons d'exploiter des indicateurs crédibilistes (fonctions de croyance: masse et communalité) pour représenter et combiner les valeurs des imperfections du système de traitement générées par les différentes sources. Seules les imperfections telles qu'une mauvaise détection, la non détection et la fausse détection ont attiré notre attention dans le cas de notre application. Pour cela, les caractéristiques de la scène telles que la taille de l'objet et la profondeur de la scène sont également utilisées pour évaluer la qualité du suivi. Cette évaluation de la qualité de la détection et du suivi des objets mobiles dans la scène permet de définir et d'allouer à chaque caméra un degré de confiance, qui est réinjecté dans la couche d'interprétation haut niveau.

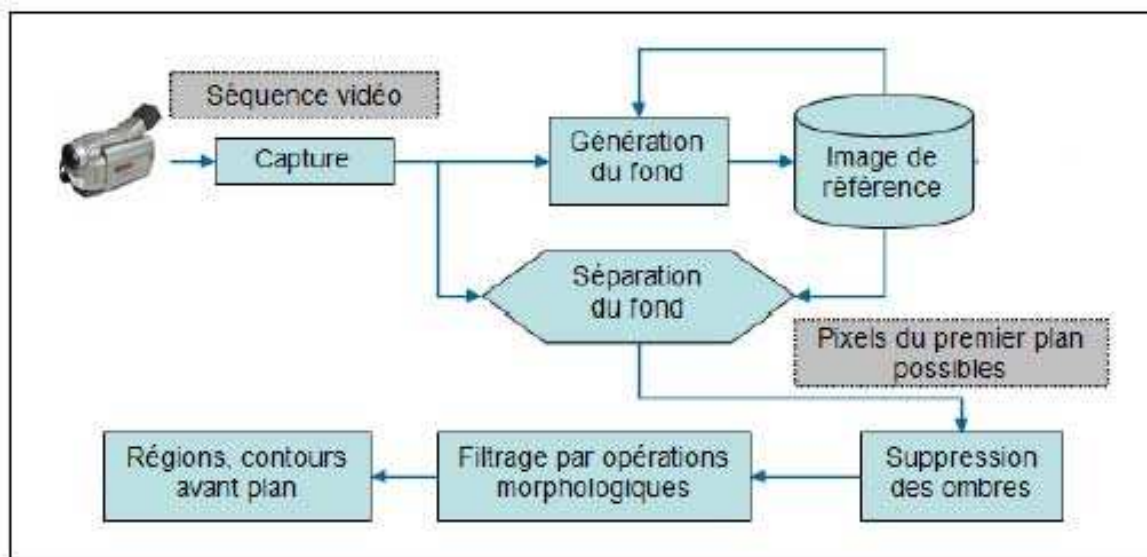


Figure 5.11. Organisation globale de l'algorithme de détection [Zia10].

5.3.2. Suivi des objets en mouvement de la scène médicale avec des indicateurs crédibilistes

Le système de suivi d'objets à partir de séquences d'images que nous exploitons est basé sur la mise en correspondance de caractéristiques entre deux images consécutives. Cette mise en correspondance est rendue difficile entre autres, par les imperfections de la détection des objets telles que la non détection et les fausses détections, une mauvaise détection et aussi la présence des occultations par exemple.

Dans notre approche, nous nous intéresserons au phénomène majeur pouvant affecter le suivi qui est lié à la qualité du processus de détection. Chaque imperfection affectant la détection des objets mobiles est caractérisée par un indicateur crédibiliste et permet d'estimer la qualité de la détection et par suite celle du suivi des objets de la scène [Mot06]. La technique que nous avons envisagée fonctionne avec le suivi de primitives d'images extraites à l'intérieur des masques des objets en mouvement. L'ensemble des données de suivis élémentaires permet de construire d'une part les trajectoires des piétons (médecins et visiteurs) puis celles des patients (mouvement du corps, des bras et des pieds) dans la scène. Au besoin, ces trajectoires peuvent être exploitées pour faire une estimation moyenne du temps de traversée par les objets dans la scène.

La mise en correspondance pour le suivi revient à rechercher le meilleur appariement entre les données d'images successives en maximisant soit une fonction de corrélation ou de ressemblance, soit de manière plus générale un critère. Il n'existe pas de technique de mise en correspondance générique. La stratégie la plus utilisée pour éviter l'explosion combinatoire de la mise en correspondance consiste à imposer des contraintes qui permettent de limiter le nombre de trajectoires candidats. Ces contraintes sont très variées et dépendent de l'objectif de l'application et des connaissances contextuelles disponibles.

La mise en correspondance d'objets pour leur suivi doit vérifier en particulier, la contrainte d'unicité. Les techniques d'association d'objets ou de primitives sont très souvent employées en vision par ordinateur. Nous les retrouvons principalement dans les domaines de la reconnaissance d'objet, de la stéréovision et de l'analyse de séquences d'images [ZSQ⁺11].

Dans le cadre d'un suivi multi-caméras, deux niveaux de suivi de manière hiérarchisée peuvent être distingués. Le premier niveau de suivi est local et est associé à chaque caméra. Il utilise l'architecture de base décrite en début de ce chapitre. Ensuite, le niveau global du suivi combine les résultats issus des différents systèmes de suivi locaux dans le repère R_0 . L'association des objets observés par les différentes caméras est réalisée en tenant compte uniquement de la ressemblance visuelle (histogramme de couleurs) et de la distance des positions des objets dans le repère R_0 . Au niveau global, pour chaque objet, la combinaison permet de nuancer l'apport de chaque suivi local en utilisant un indicateur de validation lié à chaque caméra. Pour chaque objet O_i et pour chaque caméra C_j , l'indicateur de validation λ_{ij} est réglé à son maximum si l'objet suivi est considéré comme isolé. Lorsque l'objet n'est plus observable, l'indicateur est réglé à son minimum. En présence d'occultation, mise en valeur au niveau du suivi local par la notion de groupe, l'indicateur prend la valeur du score de ressemblance w_j estimé lors de la procédure de maintien de piste au sein du groupe. La position P_i global de l'objet O_i , dans le plan R_0 , au sein du filtre global est exprimée par :

$$P^i_{global} = \frac{1}{\sum_j \lambda^i_j} \cdot \sum_j H^j \cdot (P^*_{i,j}) \cdot \lambda^i_j \quad (5.8)$$

La variable $P^*_{i,j}$ représente la position locale estimée de l'objet O_i dans l'image de la caméra C_j . La transformation homographique H_j permet de passer du repère image de la caméra C_j vers de repère de référence R_0 dans la scène. La variable λ_{ij} représente l'indicateur de validation de l'objet O_i pour le capteur C_j . La Figure 5.13 et la Figure 5.14 illustrent deux points de vue différents de la scène observée respectivement de la caméra 4 située au plafond et de la caméra 2 qui est face au lit 2.

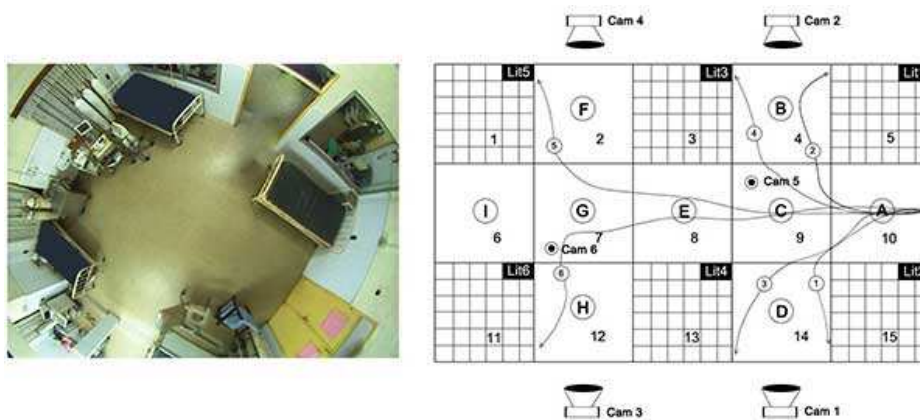


Figure 5.12. Modélisation de la salle de soins à base de régions d'intérêts [AME13].

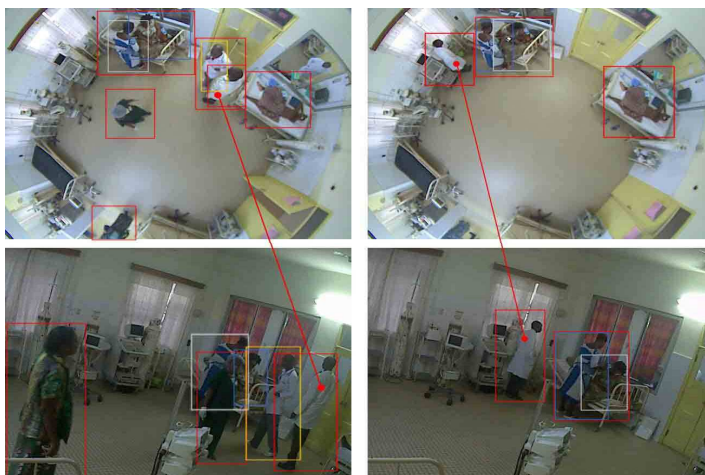


Figure 5.13. Illustration de deux points de vue différents de la scène observée respectivement de la caméra 4 et 2 à un instant t .

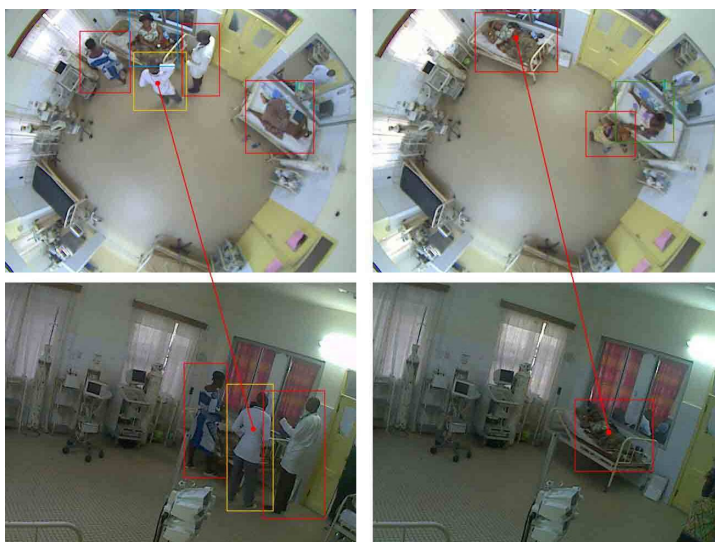
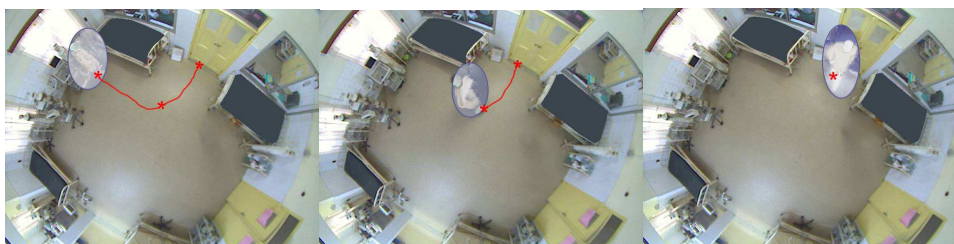


Figure 5.14. Illustration de deux points de vue différents de la scène observée respectivement de la caméra 4 et 2 à un instant $t+k$.

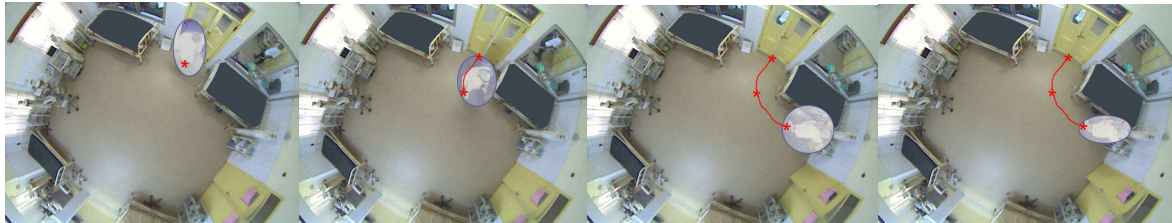
La figure 5.18, selon le cas (a), (b), (c), (d), montre les résultats de détection et de suivi d'objets mobiles «*le médecin*» dans une vidéo médicale.



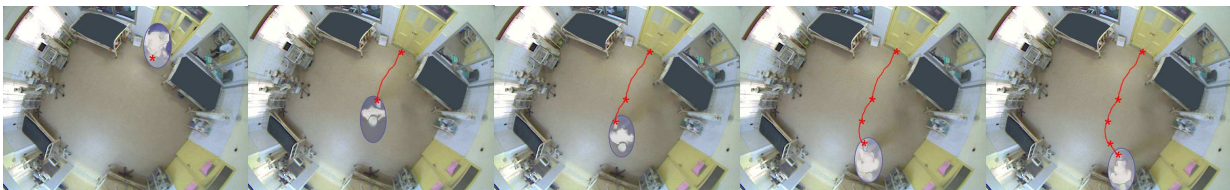
(a) Détection et suivi de l'objet mobile «*médecin*» dans le cadre de la reconnaissance du scénario "*présence du médecin au lit 1 soit pour une visite ou pour des soins*".



(b) Détection et suivi de l'objet mobile «médecin» dans le cadre de la reconnaissance du scénario "présence du médecin au lit 2 soit pour une visite ou pour des soins".



(c) Détection et suivi de l'objet mobile «médecin» dans le cadre de la reconnaissance du scénario "présence du médecin au lit 3 soit pour une visite ou pour des soins".



(d) Détection et suivi de l'objet mobile «médecin» dans le cadre de la reconnaissance du scénario "présence du médecin au lit 4 soit pour une visite ou pour des soins".

Figure 5.15. Résultats de détection et de suivi d'objets mobiles pour divers cas de scénario à reconnaître relatif à l'objet médecin.

5.4. Reconnaissance de scénarios dans des vidéos de scènes médicales

Le système de reconnaissance de scénarios dans des vidéos de scène médicales est basé sur une interprétation haut niveau de la scène par exploitation des modèles de Markov hiérarchiques cachés à durée d'état explicite dans un premier temps, puis dans un second temps, les modèles de Markov cachés hiérarchico-temporels et crédibilistes qui constituent l'apport significatif de la technique de reconnaissance proposée dans nos travaux de recherches. Ce dernier modèle, permet une gestion robuste et efficace de l'incertitude et d'imprécision des données afin d'évaluer la qualité des résultats du système de reconnaissance de scénarios et la prise en compte de l'incomplétude des connaissances médicales.

5.4.1. Extraction de caractéristiques et construction du codebook

Nous avons défini et construit un ensemble de régions d'intérêts dans l'image (le masque) afin d'identifier la position de l'objet en mouvement dans la scène observée. Chaque objet en mouvement détecté dans les images se caractérise par sa couleur, la texture de sa région d'appartenance, et ses attributs morphologiques. Ces régions d'intérêts définies dans la scène sont identifiées par des traitements des niveaux de gris, le nombre de pixels et la superficie de cette région. Chaque objet détecté est classé en exploitant ses attributs tels que sa couleur, sa texture sa taille, et sa morphologie. Ainsi, un codebook est construit (A, B, C, D, E, F, G, H, I) où chaque caractère correspond symboliquement aux différentes régions d'intérêts prédéfinies dans la vidéo. La figure 5.19 montre le résultat de la modélisation du masque de régions d'intérêt et les différentes trajectoires.

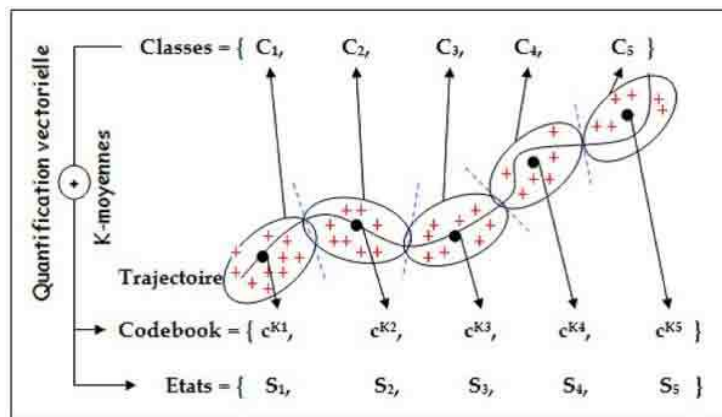


Figure 5.16. Séquence des symboles du codebook par quantification vectorielle à partir d'un exemple de trajectoire [Zia10].

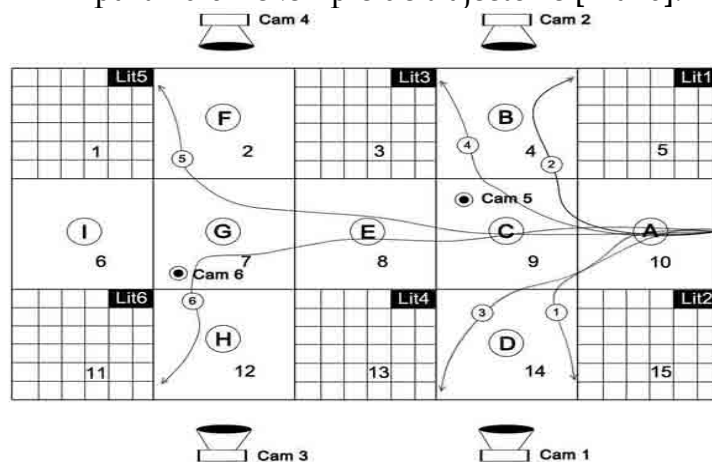


Figure 5.17. Construction d'un codebook à partir de région d'intérêts définis [AME14b].

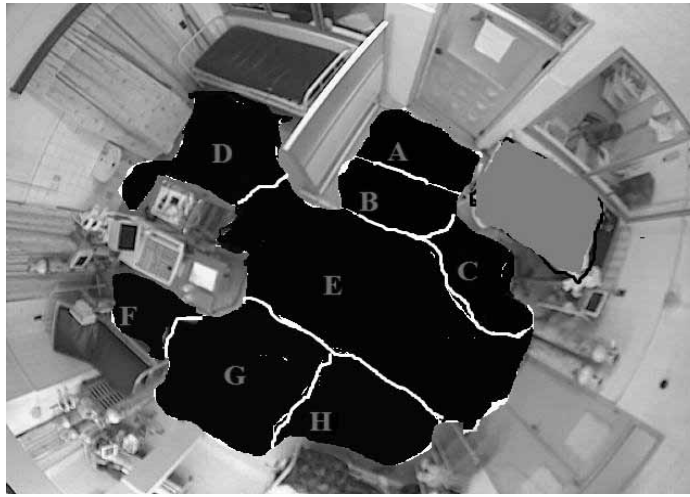


Figure 5.18. Structure et format des régions d'intérêts pour le codebook à base d'une image masque [AME14b].

5.4.2. Reconnaissance de scénarios probabilistes et crédibilistes par les modèles de Markov cachés hiérarchico-temporels

Deux principales approches ont été proposées et utilisées dans ces travaux de thèse pour reconnaître des scénarios dans des séquences vidéo médicales. Nous décrivons les expérimentations effectuées et analysons les résultats obtenus en exploitant ces approches.

5.4.2.1. Reconnaissance avec un MMC probabiliste intégrant une contrainte hiérarchique et temporelle

Les modèles de Markov cachés hiérarchiques sont une extension du modèle de Markov caché standard qui intègre une hiérarchie dans ces états cachés. Dans ces modèles, un état caché peut générer une sous-séquence d'un MMC qui est inclus dans cet état. La modélisation hiérarchique est très utile pour plusieurs applications telles que la récupération de texte, la reconnaissance manuscrite de caractères, l'indexation vidéo, et la reconnaissance de comportements.

La structure hiérarchique de ces modèles leur donne l'avantage de représenter des séquences d'observation à différents niveaux d'abstraction. Néanmoins, la hiérarchie de l'état dans le MMC hiérarchique bien que restant une idée intéressante est limitée à une structure arborescente. Malgré la capacité des modèles de Markov cachés hiérarchiques à modéliser des Sous-MMC dans un état caché du MMC père, cette structure hiérarchique qui est d'une part restrictive par rapport à certaines applications

(car arborescente), pose le problème de modélisation explicite du temps passé par le modèle dans un état donné. Dans ce modèle, l'estimation de la durée d'état du modèle reste implicite et par conséquent n'est pas adaptée à notre contexte applicatif. En effet, dans notre application, nous avons les événements de base qui composent un scénario qui sont sous forme de séries temporelles et qui font du paramètre temps un attribut de discrimination des scénarios étudiés.

Par exemple, le scénario «*le médecin visite le patient du lit 1*» se distingue du scénario «*le médecin fait des soins au patient du lit 1*» par le temps passé par le médecin au niveau du dernier état du modèle soit la présence du médecin soignant dans la région d'intérêt définie proche du lit de ce dernier patient. D'autre part, pour différencier les types de soins, une connaissance experte médicale a permis de définir un ensemble de paramètres et d'attributs temporels qui traduisent statiquement le temps moyen passé par un médecin pour une piqûre à un patient par exemple et un prélèvement de constantes matinales (*le poids, et la température, la tension artérielle du patient et etc.*). Ces deux scénarios sont discriminés tout simplement par le temps. Il convient de combiner au modèle de Markov caché hiérarchique une technique d'estimation explicite du temps passé par le modèle dans un état donné d'où l'utilisation des modèles de Markov cachés hiérarchiques et temporels.

Deux catégories de scénarios ont été testées et reconnues par le système en fonction du type d'objet «*le patient ou le médecin*» qui produit les actions que sont:

Scénarios relatifs à l'objet médecin:

- **Le *médecin visite un patient*** : par exemple la reconnaissance du scénario «*la visite du médecin au patient du lit 1*» est effectuée avec un MMC à trois états. Les états en entrée du MMC sont construits à partir de la détection et du suivi de l'objet médecin dont la présence dans l'une des régions d'intérêt est associée au symbole alphabétique prédéfini dans le codebook. Les états en sorties sont des symboles du codebook qui correspondent à un événement élémentaire et dont l'enchaînement génère le scénario important à reconnaître. Dans le cas de ce scénario, la sortie générée par le MMC est composée par les lettres A, C et B du codebook qui correspondent bien à ce scénario.

- Le *médecin fait des soins médicaux à un patient* : pour ce scénario, le même modèle de MMC utilisé dans le cas du scénario précédent «*la visite du médecin au patient du lit 1*» est utilisé pour reconnaître cette catégorie de scénario sauf qu'ici le temps est exploité comme un attribut majeur permettant de discriminer ces deux catégories de scénarios.
- Le *tour de visite du médecin à tous les patients* : ce scénario est reconnu avec un MMC à neuf états et la séquence des états en sorties de l'automate représente le scénario à reconnaître. Le tour du médecin peut commencer par n'importe quel lit, l'automate MMC exploité dans le modèle proposé à la capacité dynamique de s'adapter à la phase d'initialisation afin de reconnaître ce scénario.

La Figure 5.19 illustre la structure hiérarchique exploitée par un MMCHT pour la reconnaissance des scénarios «*le médecin visite un patient*» et «*le médecin fait des soins médicaux à un patient*». La salle de soins intensifs est constituée d'une salle de surveillance dans laquelle se trouvent les médecins surveillants et de la salle de soins proprement dite où séjournent les patients. La salle de soins comme nous l'avons modélisée précédemment dans la section plus haut est constituée de régions ou zones d'intérêts. Chaque zone d'intérêt est elle-même décomposée en événements de bases auxquels elles sont associées et permettent de générer des activités médicales avec un paramètre temporel différenciateur. En effet, la structure et le nombre d'états du MMCHT est déterminé à partir des connaissances a priori dont on dispose sur l'environnement.

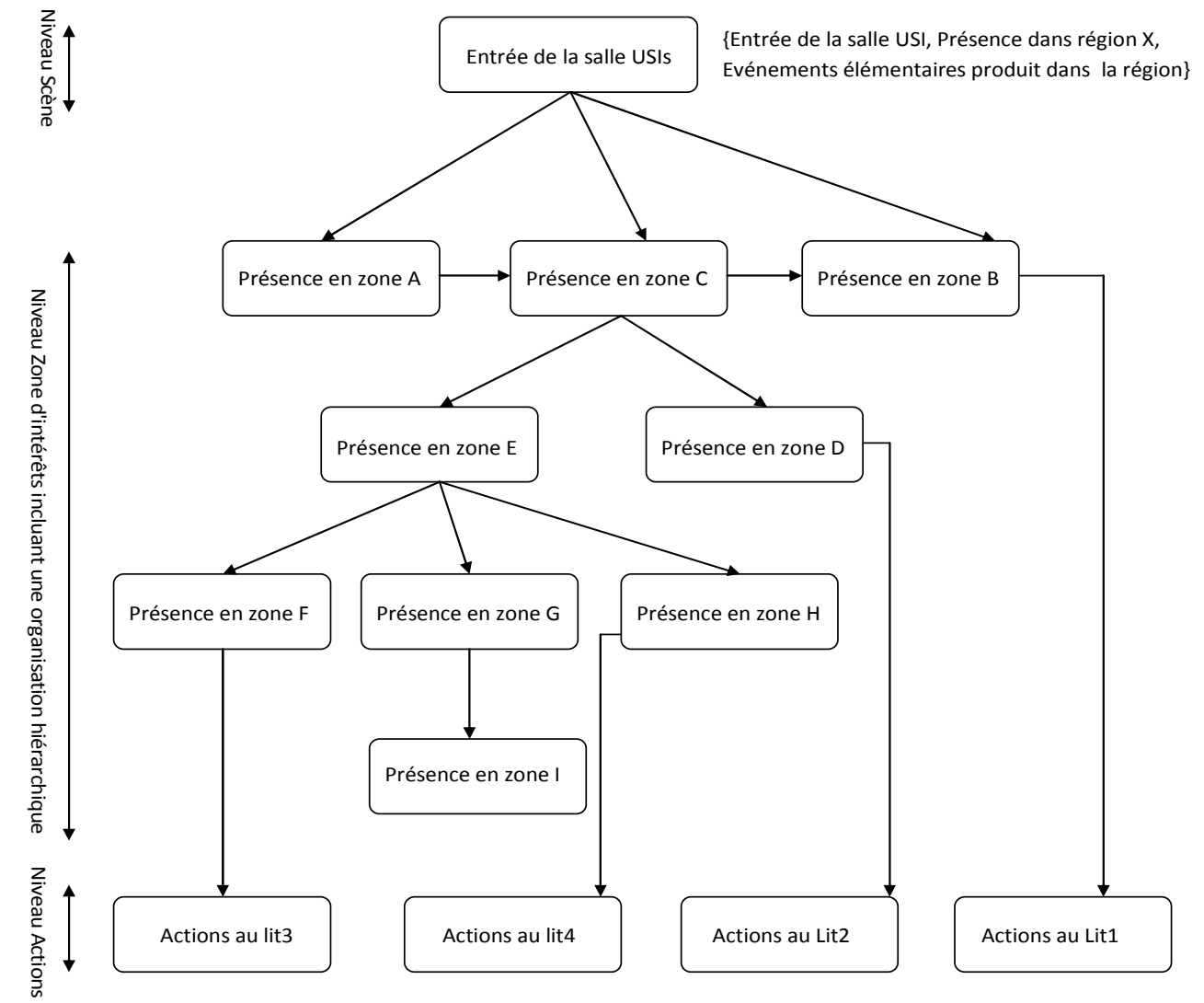


Figure 5.19. Modélisation du scénario «le médecin visite ou fait des soins médicaux à un patient d'un lit n».

La structure du MMCHT présentée ci-dessus, peut être vue comme un arbre à une seule racine au sommet de la hiérarchie. Les états émetteurs sont les feuilles et les liens entre les états sont les branches [Mer06]. Pour différencier correctement, ces deux scénarios, il a fallu modéliser explicitement la durée d'état du modèle. Nous introduisons ainsi, une modélisation explicite du temps passé par le MMC hiérarchique dans un état donné en exploitant un mélange de distributions statistiques (Poissonnière, Gaussienne). La structure du MMCHT utilisée est composée de trois principaux niveaux hiérarchiques que sont : *la racine, les zones (zones-avant et zones-arrière)* et les *actions ou événements* de base.

Au niveau "scène" : $n=1$

C'est la racine du MMCHT. Elle est constituée d'un seul état K noté R . Dans la scène étudiée, l'état R est composé de N sous-niveaux «zones d'intérêts» (par exemple, dans le cas de l'exemple du scénario étudié, $N=3$).

Au niveau "zone" : $n=2$

Ce niveau est composé de quatre sous-niveaux de zones permettant de définir les zones-avant et les zones-arrière. Le premier sous-niveau est composé de trois zones d'intérêts (zone A, zone B, zone C) dans lesquelles la présence du médecin est vérifiée. Le second sous-niveau est composé de deux zones d'intérêts (zone E et zone D) puis le troisième sous-niveau comporte aussi trois zones d'intérêts (zone F, zone G, zone H) et enfin le dernier sous-niveau de zone est associé à une seule zone d'intérêt appelée la zone I.

Au niveau "zone/actions" : $n=3$

Il est constitué, par exemple de la hiérarchie illustrée à la Figure 5.19, de quatre actions/événements de base. A chaque action est associé un événement de type présence du médecin face à un lit pour une visite médicale ou pour des soins au patient. Une action notée «*action 1*» est générée par la présence du médecin au lit 1 au préalable et correspond au fait que le médecin soit proche ou face au lit 1 du patient. En fonction du temps passé par le médecin à ce niveau (la zone d'intérêt), on peut définir un autre événement qui est le médecin fait une visite médicale ou un type de soins au patient.

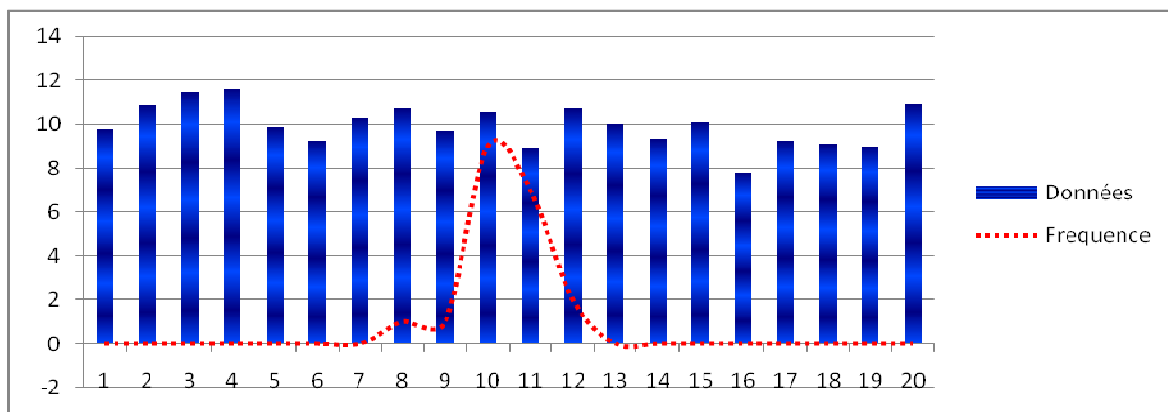


Figure 5.20. Fréquence de la durée passée par le MMC dans les événements élémentaires du scénario «*le médecin visite un patient du lit 1*».

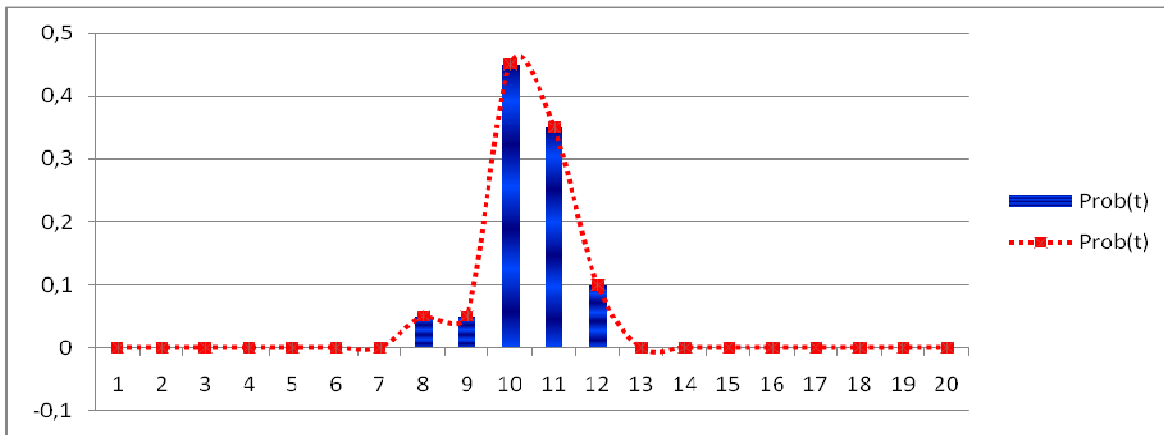


Figure 5.21. Estimation statistique de la durée passée par le modèle dans l'un des états du scénario «le médecin visite un patient du lit 1».

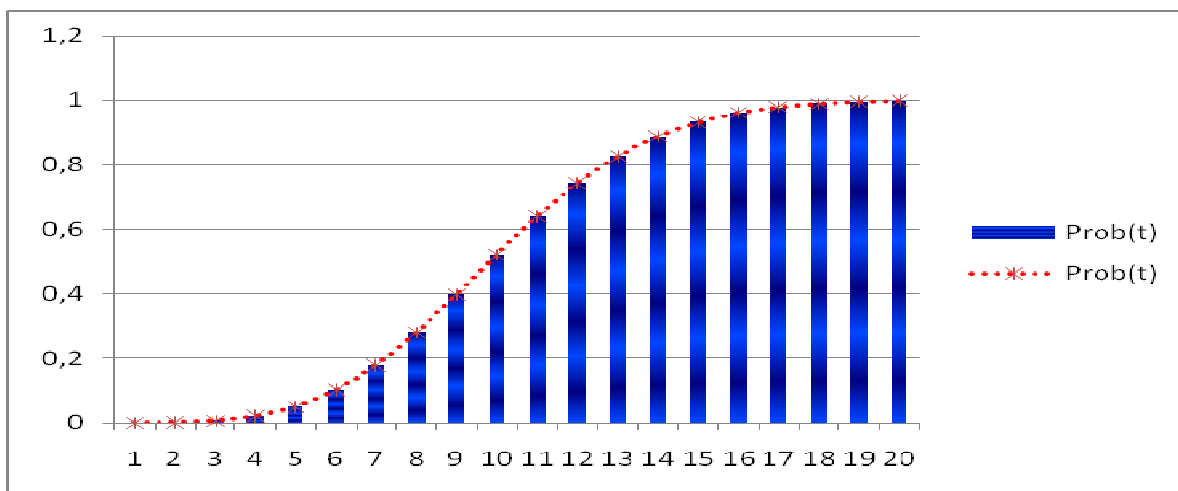


Figure 5.22. Estimation statistique de la durée passée par le modèle dans l'un des états du scénario «le médecin visite un patient du lit 1».

Les Figures 5.22 et Figures 5.23 illustrent les résultats de l'estimation statistique des paramètres de la distribution modélisant la durée du temps passé par le modèle dans un état. Les expérimentations nous ont permis de constater que l'estimation des paramètres de la loi de poisson nécessite moins de données pendant l'apprentissage tandis que la distribution gaussienne nécessite une quantité de données plus importante. L'avantage de la gaussienne est que ces paramètres sont plus stables une fois estimés sur une grande quantité de données alors que le paramètre de loi de poisson peut diverger rapidement en présence de données ne présentant pas la même ou une structure proche des données utilisées pendant la phase d'apprentissage.



(a) Présence du médecin au lit 2 pour une visite



(b) Présence du médecin au lit 4 pour une visite



(c) Présence du médecin au lit 1 pour des soins



(c) Illustration de la présence du médecin pour des soins (cas d'un relevé de constantes cliniques)

Figure 5.23. Illustration de reconnaissance de scénario liée aux activités du médecin.

Par exemple, pour la reconnaissance du scénario «*le médecin visite le patient du lit1* » et «*le médecin fait des soins médicaux au patient du lit 1*», nous avons utilisé un MMCHT à trois états dont la succession d'états à décoder après l'apprentissage est associée aux alphabets A, C et B contenus dans le codebook défini au préalable. La Figure 5.24 représente la topologie du MMCHT exploitée pour reconnaître ce type de scénario. De même, la Figure 5.25, montre l'utilisation d'un MMCHT à neuf états pour reconnaître le scénario «*le médecin fait un tour de visite globale*» dans la salle de soins intensifs pour un contrôle de l'état des patients. Les états cachés à décoder par la machine sont associés à la succession alphabétique: A-C-D-C-E-F-H-F-I.

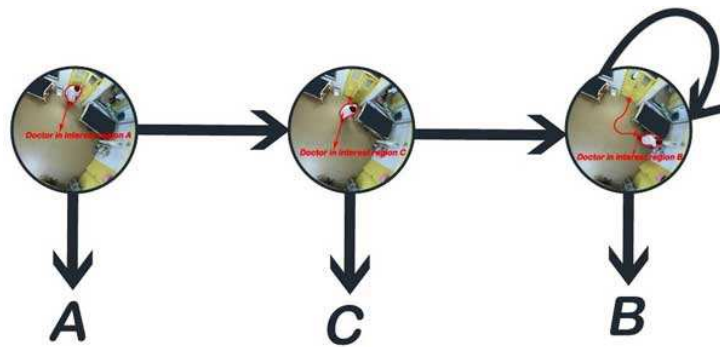


Figure 5.24. Topologie du MMCHT permettant de reconnaître le scénario «le médecin fait des soins médicaux au patient du lit 1» ou «le médecin visite le patient hospitalisé au lit1».

L'exploitation d'une distribution du temps (Poissonnière, Gaussienne,...) permettant d'explicitier le temps qui constitue, de ce fait, un paramètre important pour discriminer ces deux scénarios.

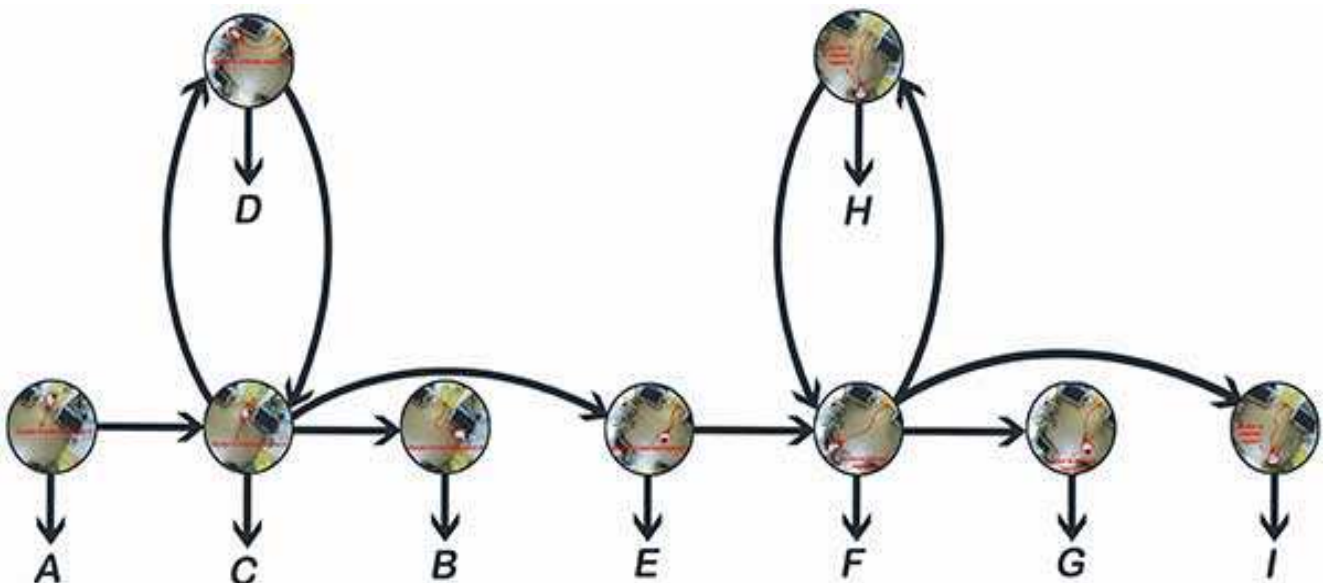


Figure 5.25. Topologie du MMCHT permettant de reconnaître le scénario «le médecin fait un tour de visite dans toutes les salles de soins» pour un contrôle ou une vérification spontanée de l'état de santé du patient.

5.4.2.2. Reconnaissance avec un MMC crédibiliste intégrant une contrainte temporelle et hiérarchique

Nous avons exploité les modèles de Markov cachés hiérarchico-temporels dans un contexte probabiliste dans la section précédente pour reconnaître des scénarios ayant une forte temporalité dans la succession des états et incluant des sous-scénarios qui peuvent être organisés sous forme arborescente. Ces modèles ont permis d'avoir des résultats très encourageants en termes de taux de reconnaissance. Cette extension des

modèles de Markov cachés a permis de s'affranchir, d'une part, des limites du modèle standard afin de résoudre efficacement le problème de gestion des sous-scénarios (branche de sous-événements «*fil*» d'un événement «*père*» de la hiérarchie). D'autre part, ce modèle permet l'estimation explicite de la variable temps qui est utilisée comme discriminant des scénarios d'une même catégorie située au même niveau de la hiérarchie. Cependant, ce modèle ne répond pas à une préoccupation majeure du domaine d'application qu'est le secteur médical.

En effet, dans un secteur tel que la médecine, qui est très spécifique avec des exigences en termes d'efficacité et de précision des résultats des traitements liés aux systèmes d'aide à la décision. Il est nécessaire de gérer les incertitudes liées aux décisions fournies par le système. Les données manipulées sont très sensibles (car en liaison avec la santé ou à la vie humaine), mais paradoxalement, il est bien connu que les connaissances dans ce domaine sont assez incomplètes et qu'on y retrouve généralement des données imprécises et incertaines à traiter. Ainsi, il est opportun de prendre en compte dans la modélisation des systèmes qui doivent traiter ces données, leurs natures et leurs caractéristiques pour espérer des résultats non obligatoirement améliorés quantitativement mais surtout qualitativement appréciables.

Fort de tout cela, nous proposons de prendre en compte et de gérer l'ensemble des imprécisions et des incertitudes résiduelles et de les propager à travers tous les traitements de niveaux supérieurs du processus de vision afin que non seulement les résultats obtenus aux couches hautes des traitements soient qualifiés, mais aussi dans la mesure du possible améliorés quantitativement.

Dans cette phase d'expérimentations, nous avons utilisé une approche de reconnaissance de scénario basée sur les outils proposés et développés dans ces travaux de recherches et qui ont été présentés dans le chapitre 4. En effet, la solution développée permet de reconnaître les activités humaines médicales dans des vidéos collectées aux USIs cardiologiques.

Une application du système ouvre des voies vers l'implémentation d'une Boîte Noire Médicale basée sur une sélection vidéo intelligente et automatique qui est une fonctionnalité du système de reconnaissance. Cette boîte noire est une base de données vidéo (multimédia) caractérisée par des scènes cliniques médicales et de comportements/états physiologiques retraçant le séjour hospitalier d'un patient aux

USIs. Le principal outil mathématique développé et exploité est le modèle de Markov caché hiérarchico-temporel et crédibiliste qui est une extension du modèle de Markov caché classique ou de base dans le cadre des Modèles de Croyances Transférables (MCT) avec gestion du temps et de la hiérarchie. L'algorithme de propagation avant-arrière (Backward and Forward) et celui de Baum-Welch ont été reformulés dans le contexte du MCT [RRP07, AME14b].

Cette extension a pour principal avantage d'une part, de profiter de la puissance et de l'efficacité des modèles de Markov confirmés par les multitudes applications existantes. Et d'autre part, elle profite de la généralité et la flexibilité des fonctions de croyances. Ce cadre générique d'exploration de données propose des outils robustes pour une prise de décision prudente sans parti-pris. Ce cadre de raisonnement permet aussi de faire une description et une modélisation très fine et plus réaliste de la nature réelle des scénarios étudiés. De plus, il offre une solution efficace pour la gestion de l'incomplétude des connaissances médicales et la modélisation de l'ignorance relative aux scénarios non encore explorés qui sont par conséquent inconnus à un instant t donné. Le doute et le conflit entre les états des différents scénarios reconnus sont aussi traités.

Dans la suite du document, nous utilisons principalement un opérateur prudent de règle conjonctive pour combiner plusieurs informations (paramètres) spécifiques issues de plusieurs sources (caméras) distinctes et indépendantes. Dans notre modèle, trois principales informations de base sont combinées dans les masses de croyances : *i) le degré de confiance alloué à la caméra, ii) la probabilité de la durée de l'état du modèle et iii) la probabilité de transition d'état.*

- i) ***le degré de confiance*** : est une variable probabiliste allouée à une caméra et dépend de plusieurs paramètres définis à partir de caractéristiques extraites des objets détectés mobiles de la scène, des paramètres intrinsèques et extrinsèques de la caméra. Ces paramètres sont estimés au cours de la phase d'acquisition des images avec un algorithme d'auto-calibration via un système de vision à perception multicaméras [AME⁺12]. La *taille* de l'objet et la *profondeur* sont utilisées avec les caractéristiques de la caméra définies plus haut pour estimer le degré de *confiance de la caméra*. Pour cela, nous avons

exploité les paramètres intrinsèques tels que la distance entre le foyer optique de l'objet détecté mobile et l'orientation de la caméra par rapport à l'objet détecté. Dans le cas des paramètres extrinsèques des caméras, nous avons utilisé le centre pixellic de l'image et sa taille focale [AME⁺12]. L'ensemble des paramètres est utilisé pour évaluer la qualité du suivi de l'objet détecté mobile par une seule caméra du système multi-caméras. La caméra optimale sélectionnée est celle qui est susceptible d'offrir la meilleure qualité de suivi de l'objet détecté mobile.

- ii) *la durée d'état probabiliste* : l'estimation de la probabilité qui représente la durée d_j passée par le MMC dans l'état j est obtenue en utilisant successivement une distribution de Poisson et une distribution Gaussienne dont la sélection est effectuée statistiquement sur la base de l'évolution temporelle des événements de base qui composent le scénario à reconnaître.
- iii) *la matrice de transition crédibiliste* : la matrice de probabilité de transition dans le cas du MMC classique est ré- estimée sous forme de fonctions de croyances. Elle est alors constituée d'éléments sous forme de masse évidentielle qui sont ensuite transformés en fonction de communalité pour des raisons de commodité et de simplicité dans le traitement des fonctions de croyances. Chacun des paramètres (*i*, *ii*, *iii*) ci dessus est combiné dans la même fonction de croyances par le biais d'un opérateur de fusion conjonctif (Equation 5.9).

$$q_1^{\Omega_t} \otimes_2 \otimes_3 (B) = q_1^{\Omega_t} (B) \cdot q_2^{\Omega_t} (B) \cdot q_3^{\Omega_t} (B) \quad (5.9)$$

Scenarios relatifs à l'objet patient :

- Le *patient s'agite* : ce scénario est modélisé et reconnu avec un MMCHT crédibiliste à quatre états. Chaque état du MMCHT modélise trois actions

(événements élémentaires): mouvement (léger et/ou saccadé) des pieds (gauche et/ou droit), mouvement (léger et/ou saccadé) des mains (gauche et/ou droit). En fonction de la succession ou de la cadence du mouvement des membres du patient combiné à ses mouvements corporels, nous pouvons décider qu'un scénario «le patient s'agite» est reconnu. Les états en entrée du MMCHT sont obtenus à partir de la détection et du suivi des mains, des pieds et du corps du patient en mouvement. Les états en sortie sont des actions de bases (événements élémentaires) telles que mouvement de la main gauche ou droite et mouvement des pieds gauche ou droite du patient à un rythme bien défini et connu. C'est à ce niveau qu'intervient l'intérêt la modélisation de la durée d'état dans le MMCHT qui combine à l'enchaînement des événements élémentaires et permet de reconnaître efficacement le scénario «*le patient s'agite*» (Figure 5.26).

- Le ***patient ressent des douleurs*** : deux MMCHTs sont conjointement utilisés pour reconnaître ce scénario. L'un est composé de quatre états dans le cadre de la reconnaissance des mouvements des membres (i.e. cas du scénario «*le patient s'agite*») et l'autre est composé de deux états afin de reconnaître les mouvements de tête (tête à gauche : « H_R » et tête à droite : « H_L »). L'exploitation conjointe (combinaison) des résultats de reconnaissance obtenus des deux MMCHTs permet de décider si «*le patient ressent des douleurs*» par le médecin. Nous avons d'autre part exploité en combinant la reconnaissance des expressions faciales du patient (reconnaissance d'émotions) à la reconnaissance obtenue avec la topologie du MMCHT de la Figure 5.27 pour améliorer la qualité de la reconnaissance du scénario «*le patient ressent des douleurs*».
- Le ***patient convulse*** : la reconnaissance de ce scénario nécessite un MMCHT de quatre états. Ce scénario diffère fondamentalement du scénario précédent par ses caractéristiques temporelles (séries temporelles liant les événements élémentaires reconnus) (Figure 5.27).

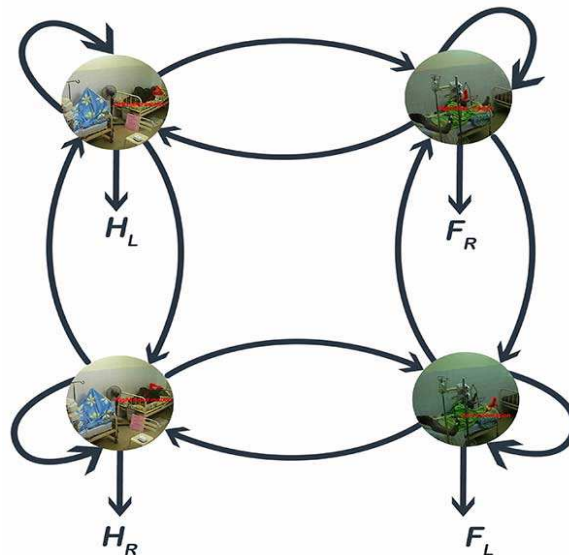


Figure 5.26. Topologie du MMC permettant de reconnaître le scénario «le patient s'agite» et le scénario «le patient ressent des douleurs».

Afin de différencier ces deux scénarios, une modélisation de la durée d'état explicite dans le MMCHT a été exploitée en introduisant une estimation probabiliste de cette durée par l'utilisation de distributions statistiques Poissonniène ou Gaussienne. L'estimation de ce paramètre temporel permet de mettre en exergue le rythme et la cadence du mouvement des membres du corps et l'ensemble du corps qui permet la détection d'une convulsion.

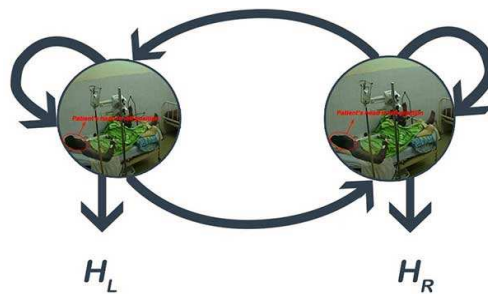


Figure 5.27. Topologie du MMCHT permettant de reconnaître le scénario «le patient convulse».

En plus de ce MMCHT, nous exploitons les résultats obtenus avec le MMCHT de la Figure 5.26 combiné aux résultats du MMCHT de la Figure 5.27 afin de détecter une convulsion du patient.

5.4.3. Résultats expérimentaux et analyse

De nombreux tests sont effectués sur notre base d'images vidéo collectées au service cardiologique à l'hôpital. Nous présentons dans cette section les résultats

expérimentaux obtenus par rapport à la qualité et le taux de la reconnaissance des scénarios mais aussi par rapport à l'apprentissage des trois scénarios testés selon leur classe d'appartenance. Cette analyse concerne non seulement la performance du système de reconnaissance automatique des scénarios mais aborde également de manière très succincte le problème complexe et difficile de l'apprentissage qui reste d'actualité et important dans le cadre de notre système de reconnaissance. Mise à part l'apport significatif du modèle hiérarchique qui permet de minimiser le nombre de MCC à utiliser pour reconnaître des scénarios et leurs sous-scénarios, deux principaux paramètres nous ont guidés dans l'analyse de nos résultats tant bien sur l'apprentissage et que sur la reconnaissance. L'impact de ces paramètres (temps et cadre de discernement) a été mis en exergue. Ces paramètres sont :

- i) le *cadre de raisonnement* qui définit le type de MMC (*probabiliste ou crédibiliste*) ;
- ii) la *distribution* utilisée pour l'estimation de la durée de l'état du MMC (*Poissonnière ou Gaussienne*).

Niveau apprentissage

Dans cette section, nous présentons les résultats expérimentaux du système d'apprentissage sur les scénarios testés et par la classe d'appartenance d'activités. S'agissant de l'apprentissage, nous avons utilisé un MMC hiérarchique dans un cadre probabiliste et puis crédibiliste avec deux variantes de distributions pour modéliser la durée d'état. Les deux variantes de distributions exploitées pour la modélisation du temps de séjour sont la distribution de poisson et la distribution gaussienne.

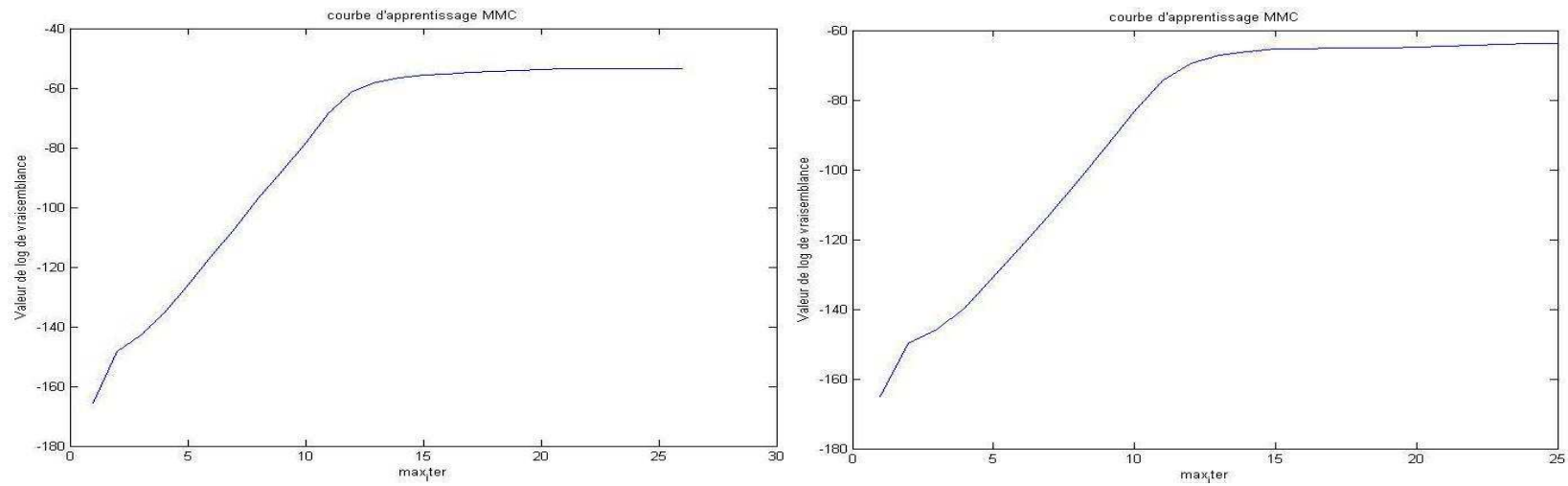
D'une part, le choix d'une distribution de type discrète comme celle de poisson est motivé par sa simplicité et sa capacité à vite converger même en présence de peu de données. De même, son exploitation dans le cadre d'une application de reconnaissance de parole [BEM06] et aussi de reconnaissance d'activités [LVW⁺04a, LVW⁺04b, DVB⁺05] ont permis d'obtenir de très bons résultats et ne nécessitent pas une grande quantité de données lors de la phase d'apprentissage avant d'être stable.

D'autre part, le choix de la distribution gaussienne pour modéliser la durée de l'état est motivée par sa capacité à obtenir des paramètres beaucoup plus stables face à

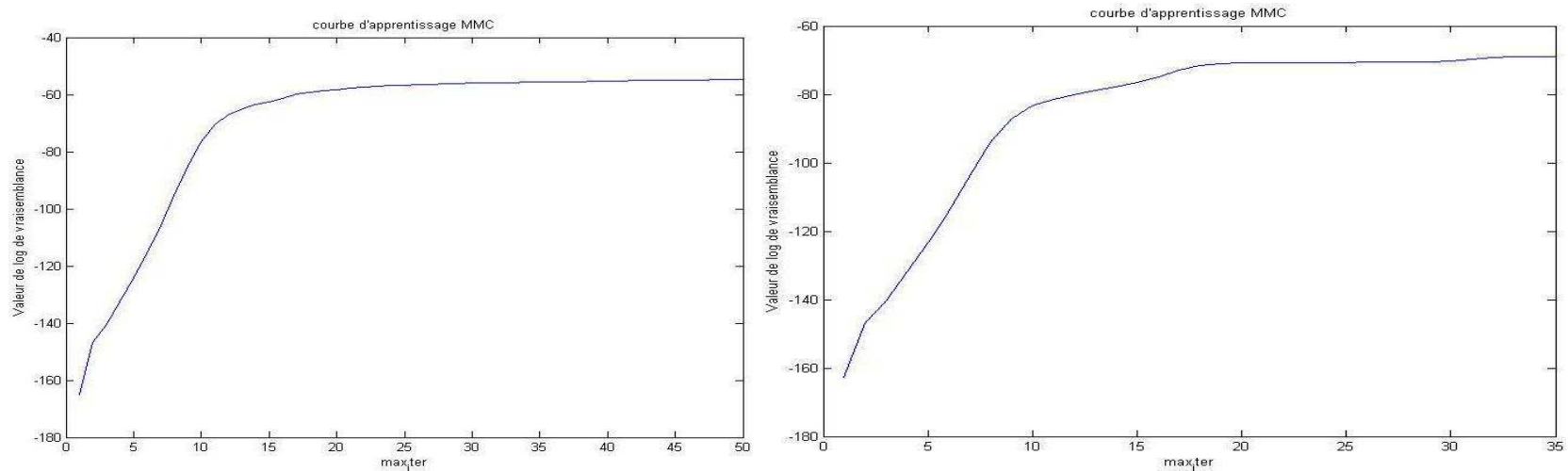
la variation des données mais cela nécessite une quantité de données plus importante. Ce type de distribution (la gaussienne positive et l'inverse) a été aussi exploité par M. Duong pour modéliser le temps de séjour des patients à l'hôpital avec de très bonnes performances [DVB+09]. A partir de nos expérimentations, nous avons obtenu la Figure 5.31 qui présente des courbes appelées «*courbe d'apprentissage*» qui est un vecteur de valeurs logarithmiques de vraisemblance (valeur ordonnée) en fonction de chaque itération (valeur en abscisse). Les itérations mettent en exergue le temps d'apprentissage en fonction de la quantité de données et le modèle du MMC utilisé. Sur ces courbes itérations, on note un temps d'apprentissage plus long avec un MMC temporel qu'un MMC classique mais ce temps d'apprentissage augmente qu'on est dans le cas d'une distribution gaussienne que par rapport à celle de poisson. En examinant, la Figure 5.31 qui montre la courbe d'apprentissage du scénario «*le médecin fait une visite médical au patient du lit1* », nous pouvons noter un temps d'apprentissage plus long du modèle standard que le modèle intégrant la contrainte temporelle tant bien dans le cas d'une distribution Poissonniène que celle Gaussienne. Cependant, ce temps d'apprentissage est encore plus importante dans le cas de la Gaussienne que le cas la Poissonniène.

La Figure 5.28 montre l'exploitation d'un temps d'apprentissage moins important qui a permis d'avoir un minimum d'itérations soit 30 itérations. Une convergence des paramètres d'apprentissage est obtenue autour d'une valeur logarithmique de vraisemblance qui vaut (60) pour un MMC probabiliste avec une loi de poisson pour la modélisation de la durée d'état. Cependant, pour le même modèle et dans le cadre crédibiliste, elle converge moins rapidement à (70) mais nécessite moins d'itérations (25) soit un temps d'apprentissage plus important. Pour une modélisation exploitant la loi gaussienne, le modèle MMC probabiliste converge autour d'une valeur logarithmique de vraisemblance qui vaut (60) tandis que celle crédibiliste converge vers une valeur de (70).

De plus, le modèle probabiliste nécessite plus de temps d'apprentissage (50 itérations avant convergence) que celle crédibiliste (35 itérations). Ces résultats viennent confirmer et conforter la capacité du cadre de raisonnement évidentielle en prenant en charge l'imperfection des données afin d'améliorer la qualité du système.



(a) Courbe de d'apprentissage d'un MMC probabiliste temporel avec une loi de poisson et (b) Courbe de d'apprentissage d'un MMC crédibiliste temporel avec une loi de poisson



(c) Courbe de d'apprentissage d'un MMC probabiliste temporel avec une gaussienne (d) Courbe de d'apprentissage d'un MMC crédibiliste temporel avec une Gaussienne

Figure 5.28. Figure Illustration de reconnaissance de scénario liée aux activités du médecin

Niveau Reconnaissance

Le système de reconnaissance proposé a été testé sur chaque scénario étudié et selon sa classe d'appartenance. Les performances du système en terme du taux de reconnaissance ont été consignées statiquement dans les Tableaux 5.2 et 5.3. De même, les Histogrammes des Figure 5.31 et Figure 5.32 montrent la distribution du taux de reconnaissance du système de reconnaissance selon chaque type de MMC que ce soit probabiliste ou crédibiliste et en fonction de la distribution qui modélise la durée d'état.

Les Tableaux 5.2 et 5.3 résument les résultats du taux moyen de reconnaissance atteint par le MMC que ce soit dans un cadre probabiliste ou crédibiliste avec une durée explicite de l'état. Ces tableaux montrent à partir des taux de reconnaissance par catégorie et type de scénario que l'utilisation d'un MMC avec la durée explicite de l'état dans le cadre crédibiliste se révèle plus efficace pour la reconnaissance relativement complexe avec des a priori incomplètes en milieu médical en comparaison à un MMC probabiliste. Le gain moyen de performance est d'environ 21,77 % pour les scénarios de la catégorie «*le médecin fait des soins médicaux au patient au lit n*» ou «*le médecin visite le patient hospitalisé au lit n*» avec distribution de Poisson avec le meilleur taux obtenu pour la reconnaissance du scénario «*le médecin ou le personnel infirmier visite le patient hospitalisé au lit1* » avec un taux qui est de 85,80 %.

La prééminence de la distribution Gamma pour la durée de l'état peut être attribuée à ses propriétés statistiques et à la pertinence des données utilisées pour l'estimation automatique de ses paramètres. Le gain moyen de performance obtenu avec la distribution Gamma est d'environ 25,86 % pour les différents lits (le meilleur taux de reconnaissance est obtenu au lit 1 et vaut 90,50 %).

Globalement, le taux de reconnaissance s'est amélioré en moyenne avec une augmentation de 30% dans le cas des MMCs crédibilistes lorsqu'on est en présence de données incomplètes par rapport aux MMCs standard probabilistes. Par contre cette évaluation est de l'ordre de 15% en moyenne lorsque les données sont assez complètes.



Figure 5.29. Quelques résultats de reconnaissance de scénarios «le patient s'agite ou ressent des douleurs».

Approches	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
MMC standard	65.00	65.00	58.00	60.60	55.00	55.00
MMCHT standard	85.00	85.00	78.00	80.80	75.80	80.60
MMCHT Crédibiliste	88.00	88.00	80.00	80.50	80.80	80.80

Tableau 5.2. Taux de reconnaissance des différentes MMCs avec une distribution de Poisson (%).

Approches	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
MMC standard	60.60	65.00	58.00	60.60	55.00	55.00
MMCHT probabiliste	90.00	90.00	80.00	90.10	85.00	80.00
MMCHT Crédibiliste	90.50	90.00	85.00	90.00	85.00	85.00

Tableau 5.3. Taux de reconnaissance des différentes MMCs avec une distribution Gaussienne (%).

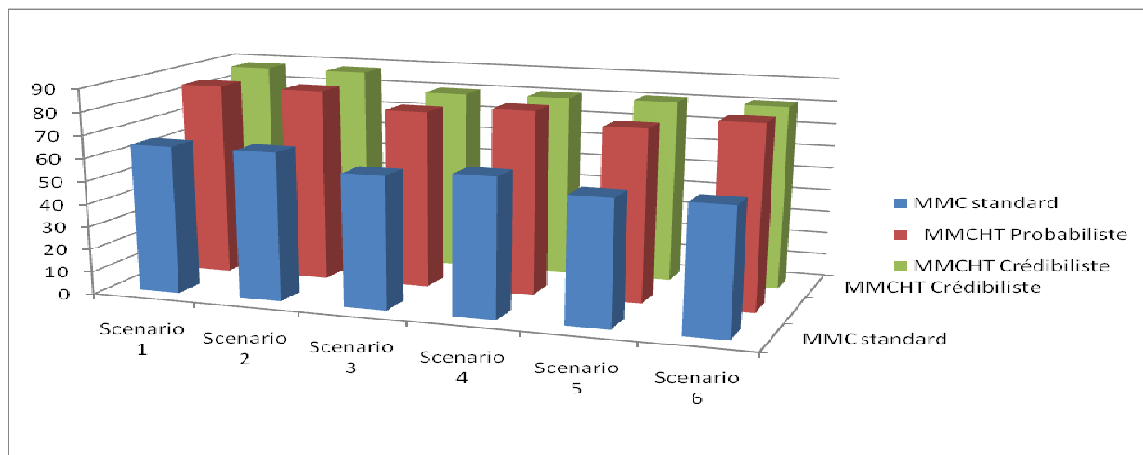


Figure 5.30. Histogramme du taux de reconnaissance des différentes MMCs avec une distribution de Poisson (%).

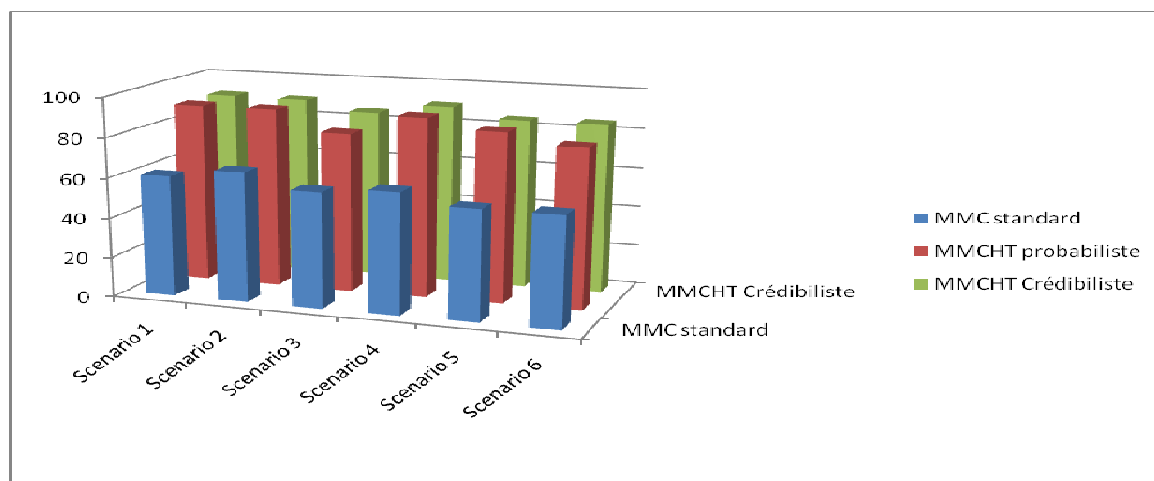


Figure 5.31. Histogramme du taux de reconnaissance des différentes MMCs avec une distribution de Gaussienne (%).

Conclusion

L'exploitation des modèles de Markov cachés représente l'une des approches les plus répandues pour la modélisation et la reconnaissance spatio-temporelle de séquences d'états. Les méthodes statistiques encore appelées techniques numériques constituent le cadre de raisonnement dans lequel se situent les modèles de Markov cachés et un atout majeur pour la robustesse de cet outil. L'extension au cadre crédibiliste permet de rendre ces outils puissants très génériques afin de prendre en charge n'importe quel type d'application avec une meilleure gestion de l'incertitude et l'imperfection des données.

Nous avons, dans un premier temps, exploité un modèle de MMC hiérarchico-temporel dans un contexte probabiliste pour reconnaître des scénarios dans des vidéos médicales. En effet, ce type de modèle permet d'hiérarchiser les structures des MMCs et de contrôler de façon explicite le temps de séjour du modèle dans un état quelconque. Ce modèle opère à deux niveaux : le niveau spatial (structure hiérarchique d'état de la séquence) puis le niveau temporel (durée d'état de la séquence). Un taux de reconnaissance assez intéressant a été obtenu avec ces modèles mais la gestion des incertitudes et des imperfections des données reste limitée considérant le domaine d'application pour lequel est destiné le système développé.

Pour gérer de façon élégante, robuste et efficace en prenant en compte l'ignorance relative à l'incomplétude des connaissances médicales avec une modélisation du doute et du conflit entre les séquences d'états, nous avons reformulé les MMCs hiérarchico-temporels dans un cadre crédibiliste. Ce cadre de raisonnement offre la possibilité de reformuler les mécanismes de base des MMCs probabilistes dans le cadre crédibiliste. Cette reformulation des mécanismes des MMCs permet de disposer d'algorithmes plus robustes et flexibles à la gestion des incertitudes et plus génériques au contexte d'applications. Les approches proposées ont été validées à travers plusieurs catégories d'expérimentations effectuées sur des données réelles afin de reconnaître des scénarios ou événements médicaux importants dans des scènes cliniques. Ces données réelles ont été collectées du système expérimental de vidéo surveillance que nous avons installé à la section cardiologique de l'hôpital (CNHU).

Conclusion générale et perspectives

L'avancée notable des résultats obtenus des travaux de recherches par les laboratoires en vision par ordinateur a permis l'intégration en temps réel de techniques plus efficaces basées sur des outils mathématiques statistiques et probabilistes dans les systèmes de vidéosurveillances de plus en plus intelligents. Dès lors, ces systèmes de vidéosurveillances sont devenus très tôt automatiques et ont réussi à aider le personnel surveillant ou de sécurité en détectant des actions suspectes et en générant des alertes informatives en temps réel, toute chose ayant facilité le développement de plusieurs applications très performantes.

1. Bilan des travaux et des contributions

Le travail de thèse développé se situe dans le cadre de la reconnaissance automatique de scénario dans des séquences vidéo. Dans ce contexte, l'élaboration du système de reconnaissance est basée sur l'analyse et l'interprétation vidéo et doit gérer l'incertitude produite par les processus de traitement du système de vision, tout ceci en prenant en compte la nature incomplète et les imperfections des données. L'objectif principal de ce travail a été de proposer une architecture organisée en couches pour un système de vidéosurveillance intelligent en milieu médical (unités de soins intensifs cardiologiques) qui intègre des fonctionnalités de reconnaissance automatique de scénarios (d'événements intéressants) dans des séquences vidéo médicales. Une synthèse des travaux effectués dans le cadre de cette thèse et de leurs résultats est présentée en trois grandes parties.

La première partie a été consacrée à une revue de littérature des différentes applications de la vidéosurveillance intelligente dans différents domaines. Cette étude a permis d'avoir une bonne connaissance des domaines d'applications avec les diverses contraintes liées à chaque type d'application. D'autre part, nous avons noté que parmi ces différentes applications, l'une des plus récentes est celle de la

télesurveillance des personnes âgées dans les maisons de retraites et encore plus récemment le domaine d'application qu'est celui médical que nous avons proposé dans ce travail. Ce qui se justifie d'autant plus que l'interprétation des vidéos médicales pour la compréhension de scènes n'avait jamais été abordée bien que dans d'autres domaines tels que le sport, les activités humaines dans les magasins, la musique et autres investigations sont en cours. Enfin, une étude du contexte d'application dans le but de cerner ses caractéristiques, ses contraintes et ses spécificités a été faite afin de les intégrer dans notre modèle.

La seconde partie du travail a été le développement d'une couche bas niveau permettant de corriger les distorsions introduites sur les données lors de la phase d'acquisition puis d'extraire, à partir des images brutes, des informations sur les objets en mouvement. Cette couche est organisée autour de deux sous-couches de traitement que sont: la sous-couche de prétraitement et la sous-couche de traitement bas niveau proprement dite. Au niveau de la première sous-couche (le prétraitement), nous avons proposé une technique robuste, élégante et moins contraignante en terme d'a priori sur le système de calibration des capteurs de vision (caméras). Cette technique est bien adaptée pour des contraintes temps réel et ne nécessite aucune connaissance a priori de la scène. Des résultats très satisfaisants en comparaison avec les techniques existantes ont été obtenus à travers les diverses expérimentations effectuées sur des données tests (image Léna et du laboratoire) exploitées par la communauté de vision par ordinateur et les données privées de la base de vidéos médicales que nous avons élaborée.

La seconde sous-couche est consacrée à l'extraction d'attributs spécifiques et morphologiques des objets d'intérêts de la scène, afin de détecter leur mouvement et de les suivre dans le temps. Pour cela, nous avons exploité les approches existantes de détection et de suivi d'objets dans des séquences d'images. Après avoir fait une analyse et dressé un bilan des différentes techniques, notre choix a porté sur un algorithme répondant à des exigences du temps réel du système d'interprétation et surtout des contraintes de gestion d'incertitude à base d'indicateurs crédibilistes pour l'estimation de la qualité de la détection et du suivi.

Enfin, la troisième partie du travail a été le cœur de notre principale contribution et concerne le développement d'une couche haut niveau d'interprétation. Cette couche d'interprétation est automatique et intègre la gestion de la structure hiérarchique et la gestion d'incertitude tant bien au niveau des données traitées que des processus de traitement de vision par ordinateur. Les couches hautes du système de reconnaissance de scénario exploitent principalement des outils statistiques dans un cadre probabiliste et crédibiliste (Modèles de Markov Cachés probabilistes et crédibilistes).

Dans une première phase, nous avons étudié les modèles de Markov cachés et leurs extensions. A partir de cette étude de l'état de l'art, nous avons proposé un modèle hybride, basé sur deux des extensions de modèles de Markov cachés, qui combine le modèle hiérarchique et celui semi-markovien. Cette approche a permis de modéliser et de reconnaître des scénarios complexes possédant une structure hiérarchisée avec une évolution temporelle de la durée d'état de chaque événement. Ce modèle est développé dans un cadre de raisonnement probabiliste et est nommé Modèle de Markov Caché Hiérarchico-Temporel (MMCHT). Dans notre approche, deux distributions probabilistes de type Poissonniène et Gaussienne ont été utilisées pour modéliser explicitement la durée d'état du modèle et par suite la durée d'un événement important (scénario).

Dans une seconde phase, face à certaines exigences du domaine d'application (médical), nous nous sommes proposé de gérer efficacement les incertitudes générées par les processus de traitement de vision. Ensuite, une prise en compte de la nature des données imprécises et l'incomplétude des connaissances qui sont des spécificités de la médecine sont intégrées dans l'approche proposée. Cela permet d'améliorer les performances de la couche haute d'interprétation ou tout au moins d'évaluer la qualité des résultats de cette dernière couche au regard de la sensibilité de l'application relative à la santé humaine. Le modèle précédemment proposé et appelé MMCHT a été reformulé dans un cadre de raisonnement crédibiliste faisant ainsi appel aux outils mathématiques de la théorie de l'évidence que sont les fonctions de croyances. Cette reformulation a donné naissance à un nouveau modèle que nous appelons Modèle de Markov Caché Hiérarchico-Temporel et Crédibiliste

(MMCHTC). L'intérêt de cette approche réside dans le fait qu'elle permet d'utiliser des indicateurs crédibilistes pour estimer à chaque étape la qualité du processus de traitement vidéo puis de propager ces incertitudes à travers les couches hautes afin d'améliorer leurs efficacités. D'autre part, elle permet de prendre en compte la nature incomplète des connaissances expertes médicales (par exemple, les médecins ne peuvent pas décrire tous les scénarios à reconnaître) à travers la modélisation de l'ignorance et du doute inter-états par les modèles de croyances transférables. Ce modèle a aussi l'avantage de gérer le conflit entre les sources d'informations que sont les capteurs de vision (caméras) et au besoin permet de combiner des informations en étant dans une attitude très prudente lors du processus de prise de décision.

2. Perspectives de recherche

Les perspectives de ces travaux de thèse concernent, d'une part, l'amélioration de la phase d'apprentissage automatique et l'approche de reconnaissance qui reste un mode supervisé. D'autre part, il est important de parfaire l'amélioration du système de surveillance actuel des patients en USIs en développant une application côté client permettant à l'utilisateur de formuler des requêtes afin de disposer d'états issus de la base de vidéos médicales pour des analyses et diagnostic de comportements des patients lors de leur séjour hospitalier. Enfin, nous développerons une idée futuriste qui offre une solution de surveillance intelligent en USIs basée sur un système multicateurs (capteurs de vision et capteurs physiologiques ou physiques) afin de combiner les différentes informations issues des deux types de capteurs pour améliorer la prise de décision.

Annexe A

Présentation des outils exploités pour le prétraitement

A1. Présentation de la géométrie épipolaire

La géométrie épipolaire, caractérisée par une matrice 3×3 singulière, décrit la relation entre deux images d'une même scène. Elle contient toutes les informations géométriques des deux images. A partir uniquement des points appariés, la géométrie épipolaire [FLT01] peut être déterminée grâce à la matrice fondamentale. Cette matrice joue un rôle important en vision par ordinateur puisqu'elle permet entre autres, sans connaissance sur la scène observée, de calculer le mouvement entre deux caméras, si leurs paramètres intrinsèques sont connus, ou d'apparier plus rapidement deux images puisque la zone de recherche d'un correspondant peut être réduite à une droite, la ligne épipolaire. La matrice fondamentale peut être estimée à partir uniquement de l'ensemble des points appariés par une méthode de corrélation en niveau de gris, que nous décrirons dans notre approche dans la suite de ce document.

A11. Définition et notions de base

La géométrie épipolaire peut être entièrement définie par deux matrices : la matrice Essentielle (E), les paramètres intrinsèques de la caméra sont connus, et par la matrice Fondamentale (F), si les paramètres sont inconnus. Ces deux matrices permettent d'associer un point d'une image à une droite dans l'autre :

- a) la matrice Essentielle E est définie dans le repère caméra (x; y)
- b) la matrice Fondamentale F est définie dans le repère image (u; v)

Le plan géométrique épipolaire est présenté à la figure A.1. Le point M, et les deux foyers optiques déterminent un plan épipolaire. L'intersection de la droite reliant les deux centres optiques avec les plans images va générer les épipôles. Ces

épipôles peuvent être dans ou hors de l'image réelle. A un point dans une image correspond une droite dans l'autre image : ceci définit la "droite épipolaire".

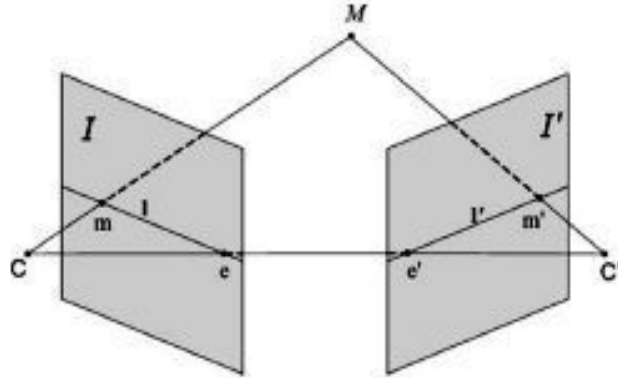


Figure A.1 : Illustration du plan géométrique épipolaire.

A12. Modélisation mathématique

Un point de la scène peut être exprimé dans les repères des deux caméras par la relation suivante :

$$M = (X, Y, Z, 1)^T \quad \text{et} \quad M' = (X', Y', Z', 1)^T \quad (\text{A.1})$$

Les deux repères sont liés par une rotation et une translation 3D, qui s'expriment par la relation matricielle suivante :

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = A * \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad \text{avec} \quad A = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{A.2})$$

Enfin, nous pouvons estimer la position du point dans le repère de droite en utilisant le système de coordonnées suivant :

$$\begin{cases} X' & = & r_{11}X + r_{12}Y + r_{13}Z + t_x \\ Y' & = & r_{21}X + r_{22}Y + r_{23}Z + t_y \\ Z' & = & r_{31}X + r_{32}Y + r_{33}Z + t_z \end{cases} \quad (\text{A.3})$$

La projection du point $m(x, y)$ dans la caméra de droite génère le point $m'(x', y')$ on peut avoir la relation de l'équation A.4 et en utilisant la relation $X=x * Z$ et $Y=y * Z$, on obtient la relation de l'équation A.4.

$$\begin{cases} x' = \frac{X'}{Y'} = \frac{r_{11} X + r_{12} Y + r_{13} Z + t_x}{r_{31} X + r_{32} Y + r_{33} Z + t_z} \\ y' = \frac{Y'}{Z'} = \frac{r_{21} X + r_{22} Y + r_{23} Z + t_y}{r_{31} X + r_{32} Y + r_{33} Z + t_z} \end{cases} \quad (\text{A.4})$$

$$\begin{cases} x' = \frac{Z (r_{11} x + r_{12} y + r_{13}) + t_x}{Z (r_{31} x + r_{32} y + r_{33}) + t_z} \\ y' = \frac{Z (r_{21} x + r_{22} y + r_{23}) + t_y}{Z (r_{31} x + r_{32} y + r_{33}) + t_z} \end{cases} \quad (\text{A.5})$$

A partir des relations mathématiques ci-dessous, nous pouvons estimer la matrice fondamentale et la matrice essentielle. Nous définissons le modèle géométrique de la caméra en tenant compte de deux transformations: une projection en perspective qui transforme un point de l'espace 3D dans le point 2D dans le repère de la caméra et une transformation de la caméra de l'image qui permet d'établir la transformation entre le repère de la caméra et l'image. La relation épipolaire peut être exprimée en coordonnées d'images. Ainsi, nous utilisons la matrice fondamentale (F), qui relie un point à son épipolaire droite:

$$l' = F * m \text{ et } l = F^T * m \quad (\text{A.6})$$

La fondamentale F est liée à la matrice essentielle E par les paramètres intrinsèques (matrice K):

$$E = K^T * F * K \text{ et } F = K^{-T} * E * K^{-1} \quad (\text{A.7})$$

Sans aucune connaissance à priori de la scène, la matrice fondamentale F peut être déterminée uniquement à partir d'une relation de correspondance entre points-images tout en faisant aussi abstraction du système de calibrage des caméras.

A13. Rectification projective et épipolaire

La rectification projective consiste à réorienter les lignes épipolaires pour qu'elles soient parallèles avec l'axe horizontal de l'image. Cette méthode est décrite par une transformation qui projette les épipôles à l'infini et dont les points correspondants sont nécessairement sur la même ordonnée. La Figure 5.7 décrit le modèle de rectification projective :

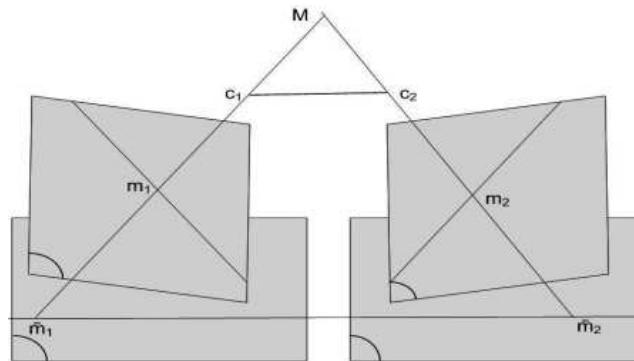


Figure 5.7 : Modèle de rectification projective.

Il existe deux types de rectification : la rectification d'images avec un système déjà calibré et la rectification d'images provenant d'un système non calibré. Dans ce document, nous nous intéressons uniquement à la rectification à partir d'images non calibrées et par conséquent à une rectification épipolaire. En se basant sur la géométrie épipolaire, la matrice fondamentale estimée et rectifiée devient une matrice fondamentale rectifiée. Il s'écrit alors sous la forme suivante :

$$F_r = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad (\text{A.8})$$

D'un point de vue purement algébrique, et à partir de la connaissance de la matrice fondamentale F et de la matrice fondamentale rectifiée F_r , nous pouvons rectifier deux images stéréoscopiques, en déterminant deux homographies H_1 et H_2 telles que [LZ99b]:

$$F = H_1^t * F_r * H_2 \quad (\text{A.9})$$

De plus, pour un ensemble de couples de points $(m_1, m_2)_k$ qui se projettent en un ensemble de points $(\tilde{m}_1, \tilde{m}_2)_k$ après rectification, la contrainte épipolaire est conservée et s'écrit :

$$m_{1_k}^t * H_1^t * F_r * H_2 * m_{2_k} = m_{1_k}^t * F * m_{2_k} = 0 \quad (\text{A.10})$$

$$\tilde{m}_{1_k}^t * F_r * \tilde{m}_{2_k} = m_{1_k}^t * H_1^t * F_r * H_2 * m_{2_k} = 0 \quad (\text{A.11})$$

A partir de la matrice fondamentale estimée, l'objectif est ensuite de déterminer deux homographies [LZ99b] H_1 et H_2 satisfaisant l'équation A.10 et l'équation A.11 qui réorientent les projections épipolaires parallèlement aux lignes des images, et par conséquent à l'axe horizontal des caméras rectifiées. Les paramètres des homographies sont estimés afin de maximiser les similitudes entre images d'origine et images rectifiées. Mais, le nombre de degrés de liberté est élevé et les homographies ne sont pas uniques. Pour ce faire, il importe alors de trouver une paire d'homographies qui réduisent le plus possible les distorsions des images rectifiées.

Annexe B

Présentation des mécanismes algorithmiques des MMCs utilisés

B1. Deux mécanismes classiques algorithmiques des MMCs utilisés

Les modèles de Markov cachés sont des automates probabilistes permettant de modéliser des données séquentielles telles que l'écriture ou la parole. Dans ces modèles, une séquence est considérée comme une suite temporelle gérée par les états inobservables (les états cachés). A chaque instant, un nouvel événement de la séquence est analysé. En effet, il s'agit d'une structure qui gère le passage d'état en état à l'aide de probabilités de transitions. En d'autres termes, chaque élément de la séquence peut être émis par un état du MMC, à l'aide de probabilités d'observations par état ce qui revient à estimer la probabilité qu'une séquence donnée soit émise par un MMC donné. Dans son tutoriel L. Rabiner [Rab89], illustre un modèle de Markov caché comme une structure composée d'états, de transitions et d'un ensemble de distribution de probabilités sur les transitions. A cette structure, on adjoint un alphabet et, pour chaque état, une probabilité d'émission des différents symboles de l'alphabet. Nous avons exploité deux mécanismes algorithmiques classiques des MMCs issus des trois problèmes principaux (apprentissage, évaluation et explication) bien connus des MMCs pour l'apprentissage des paramètres et la reconnaissance.

B11. Algorithme de Forward-Backward

Le problème de la reconnaissance peut également être vu comme un problème d'estimation de la capacité d'un modèle MMC donné à reconnaître une séquence d'observations. Ce problème se ramène aussi à un problème d'évaluation directe de la probabilité d'observation par l'équation et nécessite d'énumérer toutes les suites d'états de longueur T. Ceci implique la répétition de nombreuses multiplications et

qui augmente le niveau de complexité algorithmique. En conséquence, pour répondre à une préoccupation de réduction de la complexité algorithmique, l'algorithme d'estimation directe (*procédure forward*) pour le calcul des probabilités, et l'algorithme d'estimation de retro-propagation arrière (*procédure backward*) sont très souvent utilisés. Généralement, on considère que l'observation peut être faite en deux temps [Bau72] :

1. L'émission de la séquence d'observations $y_1^T = O(1 : t)$ et la réalisation de l'état x_i au temps t : *forward*
2. L'émission de la séquence d'observations $y_1^T = O(t+1 : T)$ et la réalisation de l'état x_i au temps t : *backward*

❖ Algorithme forward :

Supposons que dans un modèle de Markov caché λ on souhaite calculer la probabilité $P(y_1^T / \lambda)$ d'occurrence de la séquence d'observation y_1^T . La probabilité d'obtenir une séquence d'observation est en effet la somme des probabilités d'obtenir cette séquence par tous les chemins possibles et cette probabilité vaut :

$$P(y_1^T / \lambda) = \sum_{\text{chemin possibles}} \left[\left(\prod_{i=0}^{T-1} P(q_i \rightarrow q_{i+1}) \cdot P(y_{i+1} | q_{i+1}) \right) \cdot P(q_T \rightarrow q_{T+1}) \right] \quad (\text{B.1})$$

Nous pouvons calculer la probabilité d'occurrence de chacune des séquences d'observations possibles à l'aide de B1. Mais l'application directe de cette formule donne un temps de calcul de $P(y_1^T)$ et une complexité algorithmique de $O(TN^T)$. Il est préférable d'utiliser l'algorithme *Forward-Backward* [Bre00] dont le temps de calcul est en $O(TN^2)$ et pour cela on définit une variable *forward* définie par :

$$\alpha_{i,t} = P(y_1^t, x_t = i / \lambda) \quad (\text{B.2})$$

Elle exprime la probabilité d'avoir émis la séquence y_1^t tout en se trouvant dans l'état i . Cette variable peut être calculée de manière inductive :

1. Initialisation :

$$\alpha_{i,1} = P(x_1 = i) \cdot P(y_1 / x_1 = i)$$

2. Induction :

$$\alpha_{i,t} = \left(\sum_{j \in S} \alpha_{j,t-1} \cdot P(i \rightarrow j) \right) \cdot P(y_t / x_t = i)$$

On peut ainsi calculer aisément la quantité $\alpha_{i,t} = P(y_1^t, x_t = i / \lambda)$, ce qui nous amène naturellement à :

$$p(y_1^T / \lambda) = \sum_{i \in S} (\alpha_{i,T}) \quad (\text{B.3})$$

L'algorithme qui vient d'être présenté s'appelle : « *algorithme forward* ». Son nom vient du fait que l'induction est effectuée vers l'avant. Le temps de calcul avec cet algorithme est en $O(TN^2)$, alors qu'il est en $O(TN^T)$ lorsque l'on applique directement la formule de l'équation B.1.

Algorithme 1 : Procédure forward pour l'estimation de la variable α

1: **Pour** i allant de 1 à N faire : $\alpha_1(i) = \pi_i b_i(y_1)$

2: **Fin pour**

3: **Pour** t allant de 1 à $T-1$ faire :

4: **Pour** i allant de 1 à N faire: $\alpha_{t+1}(i) = \left[\sum_{j=1}^N \alpha_t(j) a_{ji} \right] b_i(y_{t+1})$

5: **Fin pour**

7: **Fin pour**

8: **Calculer** : $P(O | \lambda) = \sum_{i=1}^N \alpha_t(i)$

❖ Algorithme backward :

Il est aussi possible de réaliser le calcul à l'envers, ce qui donne l'algorithme backward. Pour cela, on définit la variable backward :

$$\beta_{i,t} = P(y_{t+1}^T / x_t = i, \lambda) \quad (\text{B.4})$$

L'estimation de cette variable $\beta_{i,t}$ suit la procédure algorithmique suivante :

1. Initialisation :

$$\beta_{i,T} = P(x_T = i).P(y_T / x_T = i)$$

2. Induction :

$$\beta_{i,t} = \left(\sum_{j \in S} P(i \rightarrow j).P(y_{t+1} / x_{t+1} = j).\beta_{j,t+1} \right)$$

On peut ainsi calculer la probabilité attendue :

$$P(y_1^T / \lambda) = \sum_{i \in S} (\beta_{i,1} * P(y_1 / x_1 = i)) \quad (\text{B.5})$$

La complexité de cet algorithme *backward* est la même que pour l'algorithme *forward* et vaut $O(TN^2)$.

Algorithme 2 : Procédure *backward* pour l'estimation de la variable β

1: **Pour** i allant de 1 à N faire : $\beta_t(i) = 1$

2: **Fin pour**

3: **Pour** t allant de 1 à $T-1$ faire :

4: **Pour** i allant de 1 à N faire: $\beta_t(i) = \sum a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)$

5: **Fin pour**

6: **Fin pour**

7: **Calculer :** $P(y_1^T / \lambda) = \sum_{i=1}^N \pi_i b_i(y_1) \beta_1(i)$

B12. Algorithme de Baum-Welch

L'algorithme de Baum-Welch est un algorithme d'apprentissage. Cette procédure forward-backward est utilisée pour construire l'algorithme de *Baum-Wech*, qui va permettre d'ajuster les paramètres du modèle afin qu'il coïncide avec le

scénario qu'il doit modéliser. Pour ce faire, à partir d'un ensemble de séquences d'observations, nous pouvons modifier les paramètres pour augmenter la probabilité. En effectuant itérativement cette modification, la probabilité atteint un maximum. On suppose les séquences d'apprentissage tirées indépendamment, on cherche à maximiser l'équation B.3. L'idée est d'augmenter la vraisemblance en mesurant les statistiques d'occurrences des observations de la séquence apprise par le modèle afin d'en déduire une ré-estimation des paramètres qui s'ajuste au mieux à cette statistique. Son but étant la maximisation de la vraisemblance d'un modèle λ , celui-ci modifie substantiellement les paramètres du modèle étudié afin d'augmenter sa vraisemblance. L'algorithme réalise son optimisation en ré-estimant les différents paramètres (A, B, Π) suivant les séquences observées. Les estimations peuvent logiquement se concevoir :

$$\begin{aligned} \overline{\pi}_i &= \text{probabilité d'être dans l'état } S_i \text{ à l'instant } t = 1 \\ \overline{a}_{ij} &= \frac{\text{nombre de transitions de l'états } S_i \text{ vers } t = 1}{\text{nombre de fois où l'on quitte } S_i} \\ \overline{b}_j(k) &= \frac{\text{nombre de fois où l'on est dans l'état } S_j \text{ en observant le symbole } v_k}{\text{nombre de fois où l'on est dans l'état } S_j} \end{aligned}$$

Il faut noter que lorsque les dénominateurs de \overline{a}_{ij} de $\overline{b}_j(k)$ sont nuls, les probabilités a_{ij} et $b_j(k)$ ne peuvent pas être calculées. Or, puisque nous réalisons des estimations sur ces coefficients, nous pouvons mettre ces probabilités à 0. Le calcul de ces coefficients est rapidement réalisé grâce aux deux matrices de paramètres α et β , délivrées par les algorithmes Forward et Backward. Après la ré-estimation des différents paramètres du modèle d'origine, l'algorithme recalcule la vraisemblance du nouveau modèle. Il va ensuite réitérer les différentes opérations de ré-estimation avec ce nouveau modèle, tant que la vraisemblance courante n'est pas maximale ($P(O/\lambda = 1)$).

Algorithme 3 : Procédure de Baum-Welch probabiliste

1: Initialisation pour le choix des paramètres du MMC (A, B, π)

2: Répéter :

3: **Pour** chaque séquence observée:

4: **Calculer** $\alpha(s_j)$ avec une procédure forward

5: **Calculer** $\beta(s_j)$ avec une procédure backward

6: **Pour** t allant de 1 à T faire

7: **Pour** j allant de 1 à N faire

8: **Calculer** $\xi(s_j)$

9: **Pour** i allant de 1 à N faire

10: **Calculer** $\gamma(s_j)$

11: **Fin pour**

12: **Fin pour**

13: **Fin pour**

14: Ré-estimer les paramètres du modèle MMC (A, B, π) avec les formules ci-dessus jusqu'à ce qu'ils deviennent stable.

15: **Tant Que** (il y a augmentation de $P(O/\lambda)$) **ou** (il y a encore des itérations à **Faire**).

Annexe C

Présentation des concepts et des mécanismes de base du Modèle de Croyances Transférables à travers un exemple didactique

Nous présentons un cas d'exemple didactique de modélisation par le MCT en nous inspirant du cas d'école : « le lancer de dé ». Notre exemple est nommé « choix automatique d'une caméra pour un service dans un réseau de caméras d'un système de vidéosurveillance ». Ce cas concerne la sélection d'une caméra détectrice d'un objet en mouvement mobile dans une vidéo afin de suivre l'objet dans un contexte de surveillance multicaméras.

Nous considérons trois caméras qui ont une vue chevauchante sur la même scène. Parmi les trois caméras (C_1 pour la caméra 1, C_2 pour la caméra 2, C_3 pour la caméra 3), une seule caméra peut être sélectionnée à la fois en fonction de sa qualité de détection. Supposons que la qualité de la détection dépend essentiellement de la position de la caméra par rapport à celle de l'objet qui met en exergue par exemple la taille de l'objet détecté mobile dans la scène. Les caméras peuvent être calibrées à trois positions possibles:

P_1 : cette position permet une *très bonne qualité de détection* si l'objet est dans une région R donnée;

P_2 : cette position permet *une bonne qualité de détection* si l'objet est dans une région R donnée;

P_3 : cette position permet une *qualité de détection assez bonne* si l'objet est dans une région R donnée.

Supposons que le système de la sélection automatique de la caméra exploite un agent pour estimer la position (P_1 , P_2 et P_3) de la caméra. Cet agent utilise un dé et le lance au hasard afin de choisir la caméra. On définit le *cadre de discernement* Ω qui est formé de l'ensemble des *hypothèses* qui représente les chiffres possibles pouvant être tirés par le dé. Par exemple, pour un dé à six faces $\Omega = \{F_1, F_2, F_3, F_4, F_5, F_6\}$ ou F_i est la face montrant le chiffre i , les hypothèses sont supposées exclusives signifiant que deux

hypothèses ne peuvent pas être vraies simultanément. Ce qui signifie par exemple que $\{F_1\} \cap \{F_2\} = \emptyset$. La face du dé tiré après le lancer est F_1 ou F_2 et la caméra 1 est sélectionnée, ensuite si la face tirée est F_3 ou F_4 la caméra 2 est sélectionnée et enfin si elle est F_5 ou F_6 la caméra 3 est choisie. L'agent est une source d'information pour le système de sélection automatique de la caméra. En se basant sur cette connaissance disponible après le lancer du dé, le système sélectionne l'une des caméras (soit la C_1 ou soit la C_2 ou la C_3) dont le champ vision est supposé détecté un objet mobile. Par exemple, on peut se poser la question de savoir quelle est la quantité de chance que l'une d'entre elle soit sélectionnée ? Comment modéliser cette connaissance fournie par l'agent au système et cette prise de décision du système ?

Dans le cadre de la théorie des probabilités, la réponse est donnée par le principe d'équiprobabilité : chaque face se voit affectée une probabilité de $1/6$ telle que la somme des probabilités soit égale à 1. Par conséquent : $P^\Omega(F_i) = \frac{1}{6}$ et par suite la probabilité n'est affectée qu'à des hypothèses singletons.

Par exemple, pour que la caméra 1 soit sélectionnée, il faut une chance égale à la probabilité $P^\Omega(C_1) = P^\Omega(F_1 \cup F_2) = \frac{1}{3}$. Il en est de même pour la caméra 2, elle vaut

$$P^\Omega(C_2) = P^\Omega(F_3 \cup F_4) = \frac{1}{3} \text{ et pour la caméra 3 elle est égale à } P^\Omega(C_3) = P^\Omega(F_5 \cup F_6) = \frac{1}{3}.$$

Soit la distribution probabiliste suivante $P^\Omega(C_1) = P^\Omega(C_2) = P^\Omega(C_3) = \frac{1}{3}$.

Il ressort qu'une situation d'équiprobabilité est la même chose qu'une situation d'ignorance sur le choix de la caméra. La valeur de la probabilité sur l'union des hypothèses du choix de la caméra 1, 2 ou 3 est implicitement déduite en faisant la somme de la probabilité sur les hypothèses singletons de laquelle on retire la partie commune. La partie commune désignant la probabilité de l'intersection ou encore du couple d'hypothèses. Pour cela, généralement, on dit qu'une probabilité est *additive*. Dans le cadre de la théorie des fonctions de croyance, la connaissance est modélisée par une distribution de masse de croyance qui différencie une connaissance d'une situation d'équiprobabilité par rapport à une situation de méconnaissance totale. En effet, dans le cas de notre exemple, pour une

méconnaissance du choix de la caméra, la totalité de la masse de croyance est affectée à la proposition formée d'une union d'hypothèses $m^\Omega(\{C_1, C_2, C_3\})=1$. Cependant, la masse de croyance allouée à cette proposition ne donne aucune information concernant la croyance sur chacune des hypothèses (choix de la caméra C1, ou C2 ou C3) composant cette proposition. C'est pour cela qu'on dit qu'une masse de croyance modélise *explicitement le doute ou l'ignorance* sur cet ensemble d'hypothèses. En comparaison à la théorie des probabilités une fonction de masse n'est pas *additive*. Ce qui signifie que la masse sur une union d'hypothèses n'est pas égale à la somme des masses des hypothèses composant l'union [KW99]. Cette propriété permet de représenter explicitement le doute entre hypothèses et constitue l'une des principales valeurs ajoutées de la théorie des fonctions de croyance par rapport à la théorie des probabilités. Outre ces valeurs ajoutées, un autre principal avantage des fonctions de croyances réside dans leur généralité qui se traduit par la modélisation des sur-ensembles des fonctions de probabilités. Les fonctions de croyance généralisent donc les fonctions de probabilités car une distribution de masses où seules les hypothèses singletons ont une masse non nulle, peut être interprétée comme une distribution de probabilités.

Annexe D

Liste des publications

1. Conférences internationales à comité de lecture

Arnaud S. R. M. Ahouandjinou, Cina Motamed, Eugène C. Ezin., Activity Recognition Based on Temporal HMM for Visual Medical Monitoring Using a Multi Camera System. *Twelfth African Conference on Research in Computer Science and Applied Mathematics, CARI 2014*: 1-12, 2014.

Arnaud S. R. M. Ahouandjinou, Cina Motamed, Eugène C. Ezin., A temporal belief hmm for human action recognition in medical videos, *11th international conference on pattern recognition and image analysis – PRIA-2013*: 492-495, 2013.

Arnaud S. R. M. Ahouandjinou, Eugène C. Ezin, Cina Motamed, Pierre Gouton: An Approach to Correcting Image Distortion by Self Calibration Stereoscopic Scene from Multiple Views. *The 8th international conference on signal image technology & internet based systems-SITIS 2012*: 389-394, 2012.

2. Journaux internationaux

Arnaud S. R. M. Ahouandjinou, Cina Motamed, Eugène C. Ezin., A temporal belief - based HMM for human action recognition in medical videos., *International journal of Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*, Accepté, pp: 1-24, vol.24, 2014.

Arnaud S. R. M. Ahouandjinou, C. Motamed, E. C. Ezin, H. Agboton, Reconnaissance de scénarios crédibilistes par les modèles de Markov cachés hiérarchico-temporels, *revue africaine de la recherche en informatique et mathématiques appliquées, ARIMA*, Soumis, 2014.

3. Posters, communications et séminaires

Arnaud S. R. M. Ahouandjinou, Cina Motamed, Eugène C. Ezin., Scenario recognition in medical environment using temporal HMM integrating Transferable Belief Model (TBM), *Posters à la 1ère Journée régionale des doctorants en automatique, JRDA-EDSPI Lille1*, 06/12/2012, France.

Arnaud S. R. M. Ahouandjinou, Patient motion recognition in medical environment (ICUs : Intensive Care Units) based on the evidential HMM using multi sensor context, Laboratoire Signal Image de Côte d'opale (LISIC) de l'Université du Littoral de la Côte d'Opale (ULCO), Salle B014, Calais, France, 14/06/2012.

Arnaud S. R. M. Ahouandjinou, Correction de distorsion de la lentille optique basée sur la géométrie épipolaire dans un réseau de caméra, *Institut de Mathématiques et des Sciences Physiques (IMSP) de l'Université d'Abomey-Calavi du Bénin*, Porto-Novo le 14/21/2011.

Arnaud S. R. M. Ahouandjinou, Reconnaissance de scénario dans une séquence d'images basée sur les Chaînes de Markov Cachée Evidentielle ou Floue : CMCE ou CMCF, *Institut de Mathématiques et des Sciences Physiques (IMSP) de l'Université d'Abomey-Calavi du Bénin*, Porto-Novo le 11/01/2011.

Bibliographie

- [AB98] P. Abecassis, et P. Batifoulier, Comment penser l'incertitude médicale à l'aide des probabilités?, *Actes du Colloque de l'Association d'Econométrie Appliquée, Econométrie de la santé XI- Maîtrise de la complexité en santé*, Lyon, France, pp. 9-10, 1998.
- [ABT10] K. Altun, B. Barshan, and O. Tunçel. Comparative study on classifying human activities with miniature inertial and magnetic sensors, *Pattern Recognition*, Vol 43, pp. 3605–3620, 2010.
- [AC09] H. Aghajan, A. Cavallaro (Eds.), *Multi-Camera Networks: Concepts and Applications*, Elsevier, pp. 335-360, 2009.
- [AEM+12] A. S. R. M. Ahouandjinou, E. C. Ezin, C. Motamed, and P. Gouton., An Approach to Correcting Image Distortion by Self Calibration Stereoscopic Scene from Multiple Views, *IEEE 8th International Conference on Signal image Technology & Internet based System*, Napples, Italy, pp. 389-394, 2012.
- [AJC11] J.R. Agustina, and G.G. Clavell, The impact of CCTV on fundamental rights and crime prevention strategies: The case of the Catalan Control Commission of Video surveillance Devices. *Computer law & security review*, Vol 27, pp. 168-74, 2011.
- [ALD11] A. Amato and V.D. Lecce, Semantic Classification of Human Behaviors in Video Surveillance Systems. *Journal WSEAS Transactions on Computers archive*. Vol 10, pp. 343-52, 2011.
- [All81] G.L Allen. A developmental perspective on the effects of "subdividing" macrospatial experience. *Journal of Experimental Psychology : Human Learning and Memory*, Vol 7, pp. 120-132, 1981.
- [AME+14] A. S. R. M. Ahouandjinou, C. Motamed, E. C. Ezin, H. Agboton, Reconnaissance de scénarios crédibilistes par les modèles de Markov cachés hiérarchico-temporels, *Journal ARIMA*, pp. 1-25, 2014.
- [AME13] A. S. R. M. Ahouandjinou, C. Motamed, E. C. Ezin, A temporal Belief-Based HMM for Human Activity Recognition in medical videos, *11th International Conference on Pattern Recognition and Image Analysis, PRIA-11-13*, Samara, Russia, pp. 492-495, 2013.
- [AME14a] A. S. R. M. Ahouandjinou, C. Motamed, E. C. Ezin, Activity Recognition Based on Temporal HMM for Visual Medical Monitoring Using a Multi-Camera System, *12ème Colloque Africain pour la Recherche en Informatique et en Mathématiques Appliquées CARI'14*, Sénégal, pp. 1-12, 2014.
- [AME14b] A. S. R. M. Ahouandjinou, C. Motamed, E. C. Ezin, A temporal Belief-Based HMM for Human Activity Recognition in medical videos, *in international journal "Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*, Ed. Springer, pp. 1-25, 2014.
- [ASA+00] K.A. Al-Shalfan, J.G.B. Haigh and S.S. Ipson, Direct algorithm for rectifying pairs of uncalibrated images, *Electronics Letters* 36 (5), pp. 419–420, 2000.

- [Avi03] P. Avis, Surveillance and Canadian maritime domestic security, *Canad. Military J.*, pp. 9–15, 2003.
- [Bati94] Batifoulier, P., Incertitude sanitaire et risque probabilisable, *Journal d'Economie Médicale*, Vol 2, pp. 3-17, 1994.
- [Bau72] L.E. Baum. An inequality an associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, vol3, pp. 1-8, 1972.
- [BEM06] A. Benouareth, A. Ennaji, and M. Sellami, HMMs with Explicit State Duration Applied to Handwritten Arabic Word Recognition, *International Conference on Pattern Recognition ICPR*, Hong Kong, Chine, pp. 897-900, 2006.
- [BF95] Y. Bengio and P. Frasconi, Input/Output HMMs for sequence processing, *IEEE Transactions on Neural Networks*, pp. 519-523, 1995.
- [BG95] H. Buxton and S. Gong. Advanced visual surveillance using Bayesian Networks, *In International Conference on Computer Vision*, Cambridge, Massachusetts, USA, pp. 111-123, 1995.
- [Bla86] S.S. Blackman, Multiple-target tracking with radar applications, *Artech House, Inc Boston*, pp. 401-409, 1986.
- [Blo96] I. Bloch, Information combination operators for data fusion: a comparative review with classification. *IEEE Trans. Syst. Man Cybern.* 6, pp. 52–67, 1996.
- [BM98] F. Bremond and G. Medioni. Scenario Recognition in Airborne Video Imagery. *Interpretation of Visual Motion Workshop, Computer Vision and Pattern Recognition (CVPR98)*, Santa Barbara, USA, pp. 211-216, 1998.
- [BMN96] B. Bouchon-Meunier, H. T. Nguyen, Les incertitudes dans les systèmes intelligents, *Que Sais Je? Presse Universitaire de France*, 1996.
- [Bor00] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, A Comparison of Probabilistic, Possibilistic and Evidence Theoretic Fusion Schemes for Active Object Recognition. *Springer-Verlag, Computing* 62, pp. 293–319, 2000.
- [Bou00] J.Y. Bouguet, Calibration Toolbox: <http://www.vision.caltech.edu/bouguet/>, 2000.
- [Bouc95] B. Bouchon-Meunier, La logique floue et ses applications. *Vie artificielle. Addison-Wesley*, pp. 1-257, 1995.
- [BPV04] H.H. Bui, Q. Phung, and S. Venkatesh., Hierarchical Hidden Markov Models with General State Hierarchy, *In Proceedings of the Nineteenth National Conference on Artificial Intelligence*, San Jose, California, pp. 324–329, 2004.
- [BSF88] Y. Bar-Shalom, T.E. Fortmann, Tracking and data association, *Academic Press, Orlando, Fla, USA*, pp. 1-353, 1988.
- [BSL95] Y. Bar-Shalom, X. Li, Multitarget-Multisensor Tracking: Principles and Techniques, *YBS Publishing*, pp. 8-37, 1995.
- [Bug07] A. Bugeau, détection et suivi d'objets en mouvement dans des scènes complexes, application à la surveillance des conducteurs, *Phd thesis, Université de Rennes 1, France*, 2007.

- [BVW01] H. H. Bui, S. Venkatesh, and G.A.W. West, Tracking and surveillance in wide-area spatial environments using the abstract hidden markov model, *Int. Journal Pattern Recognit. Anal. Intell.*, Vol 15, pp. 177-195, 2001.
- [BVW02] H. H. Bui, S. Venkatesh, and G. West. Policy Recognition in the Abstract Hidden Markov Model, *Journal of Artificial Intelligence Research* 17, pp. 451-499, 2002.
- [BXL11] Y.-W. Bai, Z.-L., Xie and Li, Z.-H. Design and Implementation of a Home Embedded Surveillance System with Ultra-Low Alert Power. *IEEE Transactions on Consumer Electronics*. 75, pp. 153-159, 2011.
- [Cat03] M. Cattaneo. Combining belief functions issued from dependent sources. In *Proc. of the Third Int. Symposium on Imprecise Probabilities and Their Applications*, Carleton Scientific., Lugano, Switzerland, pp. 133-147, 2003.
- [CCT96] L. Chaudron C. Castel and C. Tessier. 1st Order CCubes for the Interpretation of Petri Nets : an Application to Dynamic Scene Understanding. *TAI'96 8th International conference on tools with artificial intelli-gence*, Toulouse, France, pp. 366-373, 1996.
- [CDG93] P. Gaborit C. Dousson and M. Ghallab. Situation Recognition : Representation and Algorithms. In *International Joint Conference on Artificial Intelligence (IJCAI)*, Chambéry, France, pp. 166-172, 1993.
- [CFA13] F. F. Chamasemani and L. S. Affendey, Systematic Review and Classification on Video Surveillance Systems, *I.J. Information Technology and Computer Science*, 07, pp. 87-102, 2013.
- [CFP+10] D. Conte, P. Foggia, G. Percannella, F. Tufano and M., Vento, A Method for Counting Moving People in Video Surveillance Videos. *EURASIP Journal on Advances in Signal Processing*. pp. 1-10, 2010.
- [CGP+04] R. Cucchiara, C. Grana, A. Patri, G. Tardini, and R. Vezzani, Using computer vision techniques for dangerous situation detection in domotic applications. *Proc. IEE Workshop on Intelligent Distributed Surveillance Systems*, London, Royaume-Uni, pp. 1-5, 2004.
- [CH96] I.J Cox , S.L. Hingorani, An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Propose of Visaul Traking, In *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 18, pp. 138-150, 1996.
- [CLL07] C.L. Chiang, C.C. Lien and C.H. Lee, Scene-based event detection for baseball videos. *Journal of Visual Communication and Image Representation*, vol. 18, no. 1, pp. 1-14, 2007.
- [DB01] M. Dojat and L. Brochard, Knowledge-based Systems for Automatic Ventilatory Management, *Saunders: Philadelphia*, pp. 379-396, 2001.
- [Dem67] A.P. Dempster. Upper and lower probabilities induced by multiple valued mappings. *Annals of Mathematical Statistics*, pp. 325-339, 1967.
- [Dem68] A.P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society*, pp. 205-247, 1968.
- [Den04] T. Denoeux. Théorie des fonctions de croyance et classification, *Rapport d'activité, Séminaire, thème ASTRID, UTC*, 2004.

- [Den08] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence, *In Artificial Intelligence*, pp. 234-264, 2008.
- [DG95] N. Dimitrova and F. Golshani. Motion recovery for video content classification. *ACM Trans. Inf. Syst.*, pp. 408-439, 1995.
- [DH07] M. Duong and D. He. A segmental hidden semi-Markov model (HSMM) based diagnostics and prognostics framework and methodology, *Mech. Syst. Signal Processing*, pp. 2248–2266, 2007.
- [DM04] T. Denoeux and M.H. Masson. EVCLUS: evidential clustering of proximity data. *IEEE Trans. Systems, Man and Cybernetics B*, Hyderabad, India, pp. 95-109, 2004.
- [DM07] C. Dousson, and P., L. Maigat, Chronicle Recognition Improvement Using Temporal Focusing and Hierarchization, *IJCAI 2007*, pp. 324-329, 2007.
- [Dou94] C. Dousson. Suivi d'évolution et reconnaissance de chroniques. *Phd thesis, Universite Paul Sabatier*, Toulouse, France, 1994.
- [DP86a] D. Dubois and H. Prade, The principle of minimum specificity as a basis for evidential reasoning. *In Uncertainty in knowledge-based systems*, Springer-Verlag, vol 12, pp. 75–84, 1986.
- [DP86b] D. Dubois and H. Prade, A set-theoretic view of belief functions: Logical operations and approximations by fuzzy sets. *Int. Jour. of General Systems*, Vol 12, pp. 193–226, 1986.
- [DP87] D. Dubois and H. Prade. Fuzzy numbers: An overview. *In J.C. Bezdek, editor, Analysis of Fuzzy Information, Vol.I: Mathematics and Logic*, pp. 3-39, 1987.
- [DP87c] D. Dubois and H. Prade, Théorie des possibilités, applications à la représentation des connaissances en informatique, Masson, Paris, 1987.
- [DP87d] D. Dubois and H. Prade, Une approche ensembliste de la combinaison d'informations incertaines. *Revue d'intelligence artificielle*, Vol 1 N°4, pp. 23-42, 1987.
- [DRH11] C. L. Devasena, R. Revathi, M. Hemalatha, Video Surveillance Systems – A Survey, *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 4, No 1, pp. 635-642, 2011.
- [DS04] F. Delmotte and P. Smets. Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 34, pp. 457-471, 2004.
- [DVB+05] M. Duong, T. Vu., H.H., Bui, H., Phung, Q., Dinh and V., Svetha, Activity recognition and abnormality detection with the switching hidden semi-Markov model, *Proceedings of the IEEE Computer Society Conference, CVPR*, pp. 838-845, 2005.
- [DVB+09] M. Duong, T. Vu., H.H. Bui, H. Phung, Q. Dinh and V. Svetha Efficient duration and hierarchical modeling for human activity recognition, *Artificial intelligence*, vol. 173, no.7-8, Elsevier BV, Amsterdam, Netherlands, pp. 830-856, 2009.
- [EEG01] A. Esposito, E. C. Ezin and C. A. R. García, Speech Noise Cancellation Based on a Neuro-Fuzzy System: Further Improvements. *Journal of*

- Advanced Computational Intelligence and Intelligent Informatics JACIII*, Vol 5(1), pp. 15-21, 2001.
- [FBG99] W. Friesdorf, B. Buss and M. Gobel: Monitoring alarms-the key to patient's safety in the icu? *Intensive Care Medecine*, Vol 25, N°12, pp. 1350-1352, 1999.
- [FLT01] O. Faugeras, Q. T. Luong, the Geometry of Multiple Images, *MIT Press*, Cambridge, MA, 2001.
- [FPK05] R. Dahyot F. Pitié, S. Berrani and A. Kokaram. Online multiple object tracking using candidate selection and the viterbi algorithm. *Proc. Int. Conf. Image Processing*, Geova Italy, pp. 1-8, 2005.
- [FRP11] R. Fonseca, and P. P. Rodrigues, the importance of system integration in intensive care units: A review, *International Conference on Health Informatics, HEALTHINF*, pp. 142-147, 2011.
- [FST98] S. Fine, Y. Singer and N. Tishby. The Hierarchical Hidden Markov Model: Analysis and Applications, *Machine Learning*, Vol 32,N°1, pp.41-62, 1998.
- [Fus00] A. Fusiello, Epipolar Rectification, 2000. <http://www.sci.inivr.it/~fusiello>
- [GB00] Geradts, Z., and Bijhold, J.: Forensic video investigation, in *Foresti, G.L., Mahonen, P., and Regazzoni, C.S. (Eds.): Multimedia video based surveillance systems (Kluwer Academic Publishers)*, Boston, pp. 3-12, 2000.
- [GCB+05] V. Girondel, A. Caplier, L. Bonnaud, and M. Rombaut. Belief theory-based classifiers comparison for static human body postures recognition in video. *International Journal of Signal Processing*, pp. 29-33, 2005.
- [GD00] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems. In *Proc. Europ. Conf. on Computer Vision, Dublin, Ireland*, pp. 446-461, 2000.
- [Gel98] M. Gelgon, Segmentation spatio-temporelle et suivi dans une séquence d'images : application à la structuration et à l'indexation de vidéo ». Thèse de doctorat, Université de Rennes 1, France, 1998
- [Gha96] Malik Ghallab. On Chronicles: Representation, On-line Recognition and Learning. *5th International Conference on Principles of Knowledge Representation and Reasoning (KR'96)*, Cambridge, USA, pp. 597-606, 1996.
- [GJH01] A. Galata, N. Johnson, and D. Hogg, Learning variable length Markov models of behaviour, *Int. Journal of Comp. Vision and Image Understanding*, 81(3), 2001, pp. 398-413, 2001.
- [GM09] M. Görges and B. A. Markewitz, D.R. Westenskow, Improving alarm performance in the medical intensive care unit using delays and clinical context, *Anesth Analg*, pp. 1546-1552, 2009.
- [GNS02] R. Gerber, H. Nagel and H. Schreiber. Deriving Textual Descriptions of Road Traffic Queues from Video Sequences, *The 15-th European Conference on Artificial Intelligence (ECAI)*, Lyon, France, pp. 736-740, 2002.
- [GRJ02] Greenhill, D., Remagnino, P., and Jones, G.A.: VIGILANT: contentquerying of video surveillance streams, in *Remagnino, P., Jones, G.A., Paragios, N., and Regazzoni, C.S. (Eds.): Video-based Surveillance Systems (Kluwer Academic Publishers, Boston, USA, pp. 193-205, 2002.*

- [Guy07] T. Guyet, *Interprétation collaborative des séries temporelles multi variées. Application à des données de réanimation médicale, Phd Thesis Institut National Polytechnique de Grenoble, France, 2007.*
- [Har00] D. Harmanec. Measures of uncertainty and information. *International Journal of Approximate Reasoning*, pp. 365-390, 2010.
- [HBN00] S. Hongeng, F. Bremond and R. Nevatia, Bayesian Framework for Video Surveillance Application, *ICPR00, Barcelona, Spain, pp.164-170, 2000.*
- [HCR05] Z. Hammal, A. Caplier, and M. Rombaut. Belief theory applied to facial expressions classification. *In Int. Conf. on Advances in Pattern Recognition, Bath, United Kingdom, pp.1-9, 2005.*
- [Hei11] E. Heilmann, Video surveillance and security policy in France: From regulation to widespread acceptance. *Information Polity*. 16, pp. 369-77, 2011.
- [HHD00] Haritaoglu, I., Harwood, D., and Davis, L.S.: W4: real-time surveillance of people and their activities, *IEEE Trans. Pattern Anal. Mach. Intell.*, 22, (8), pp. 809-830, 2000.
- [HIZ03] R.I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, second ed., Cambridge University Press, 2003.
- [HJB00] R.J. Howarth and H. Buxton. Conceptual Descriptions from Monitoring et Watching Image Sequences, *Image and Vision Computing*, 18, pp. 105-135, 2000.
- [HJB02] A.J., Howell and H. Buxton, RBF Network Methods for Face Detection and Attentional Frames, *Neural Processing Letters*, pp.197-211, 2002.
- [HJB98] A. J. Howell and H. Buxton. Learning Identity with Radial Basis Function Networks, *Neurocomputing*, Vol. 20, pp. 15-34, 1998.
- [HLT07] H. He, H. Li, and J. Tan. Real-time daily activity classification with wireless sensor networks using hidden markov model, *In 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Los Amiltos, USA, pp. 3192 -3195, 2007.*
- [HN01] S. Hongeng and R. Nevatia. Multi-agent event recognition, *Proc. of the Int. Conf. on Computer Vision, ICCV'01, Vancouver, Island, pp. 84-93, 2001.*
- [HNB04] S. Hongeng, R. Nevatia, and F. Bremond. Video-based event recognition and probabilistic recognition methods. *Computer Vision and Image Undertanding*, Vol 96, pp. 129-162, 2004.
- [Hon02] S. Hongeng. A Unified Bayesian and Logical Approach for Video-base Event Recognition. *PhD thesis, University of Southern California, USA, 2002.*
- [HTW+04] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, man and cybernetics C*, 34(3), pp. 334-352, 2004.
- [IB00] Y.A Ivanov and A.F Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol 22(8), pp. 852-872, 2000.

- [JGH⁺10] J. Jeong, Y. Gu, T. He, D.H.C. Du, Virtual Scanning Algorithm for Road Network Surveillance. *IEEE Transactions On Parallel And Distributed Systems*. 21, pp. 1734-49, 2010.
- [JMC08] J-M. Jolion, C. Leignel, Projet ANR CAnADA «Comportements Anormaux Analyse Détection Alerte» Étude bibliographique suivi de personnes et trajectoire dans un réseau de caméras, *LIRIS, UMR 5205 CNRS, INSA de Lyon, Université Claude Bernard Lyon 1, France*, 2008
- [KAA86] H.A. Kautz and J. F. Allen, Generalized plan recognition. *In proceedings of the 5th National Conference on Artificial Intelligence*, Philadelphia, USA, pp. 32-37, 1986.
- [KBS00] S.B. Kang. Catadioptric self-calibration. *In CVPR00*, pp. 201-207, 2000.
- [KHM⁺00] J. Krumm, S. Harris, B. Meyers, B. Brumit, M. Hale, and Shafer, S, Multi-camera multi-person tracking for easy living. *Third IEEE Int.Workshop on Visual Surveillance*, Ireland, pp. 8-11, 2000.
- [Kij03] E. Kijak. Structuration multimodale des videos de sport par models stochastiques. *Thèse de doctorat, Université de Rennes 1, France*, 2003.
- [Kon00] J. Konrad. Handbook of image and video processing. *Academic press*, pp. 207- 225, 2000.
- [KRD⁺06] A. Kokaram, N. Rea, R. Dahyot, M. Tekalp, P. Bouthemy, P. Gros and I. Sezan. Browsing sports video (Trends in sports-related indexing and retrieval work), *IEEE Signal Processing Magazine*, pp. 47-58, 2006.
- [KS86] R. Kowalski and M. Sergot, A logic-based calculus of events. *New Generation Computing*, Vol.4, pp. 67-95, 1986.
- [Kun04] L.I. Kuncheva. Classifier ensembles for changing environments, *In Int. Workshop on Multiple Classifier Systems*, volume 3077, Cagliari, Italy, pp. 1-15, 2004.
- [KW99] G.J. Klir and M.J. Wierman, Uncertainty-based information, Elements of generalized information theory, 2nd edition. *Studies in fuzzyness and soft computing. Physica-Verlag*, 1999.
- [LCT07] C. Lesire-Cabaniols and C. Tessier, Particle Petri net-based estimation in hybrid systems to detect inconsistencies. *In DCDS'07, 1st IFAC Workshop on Dependable Control of Discrete Systems*, Paris, France, pp. 1-6, 2007.
- [LCV02] E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief function combination and conflict management. *Information Fusion*, Vol 3, pp. 149-162, 2002.
- [Lev86] S. E. Levinson, Continuously Variable Duration Hidden Markov Models for Automatic Speech Recognition. *Computer, Speech and Language* pp. 29-45, 1986.
- [LJK11] M.J.H. Loomans, C. J.Koeleman, With, P.H.N.d. Low-Complexity Wavelet-Based Scalable Image & Video Coding for Home-Use Surveillance. *IEEE Transactions on Consumer Electronics*, pp. 507-15, 2011.
- [LM03] J. L. Lisani et Jean-Michel Morel. Detection of major changes in satellite images. *In IEEE International Conference on Image Processing, Barcelona, Spans*, pp. 941-944, September, 2003.

- [LM11] M.J. Leotta, and J.L. Mundy, Vehicle Surveillance with a Generic, Adaptive, 3D Vehicle Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol 33, pp. 57-69, 2011.
- [LM98] L. Moreno, Factorial HMMs for acoustic modeling, *ICASSP*, Washington, USA, pp. 813-816, 1998.
- [Luv10] B. Luvison., Détection non supervisée d'évènements rares dans un flot vidéo: Application à la surveillance d'espaces publics, Université Blaise Pascal - Clermont II, *Phd thesis*, 2010.
- [LVW+04a] S. Luhr, S. Venkatesh, G. A. West and H. H. Bui, Duration abnormality detection in sequences of human activity, *Technical report Curtin University of Technology*, Perth, 2004.
- [LVW+04b] S. Luhr, S. Venkatesh, G. A. West and H. H. Bui, Explicit state duration HMM for abnormality detection in sequences of human activity, *PRICAI*, Auckland, New Zealand, pp. 983-984, 2004.
- [LWJ11] X. Luo, Y. Wu, Y. Huang and J. Zhang, Vehicle flow detection in real-time airborne traffic surveillance system. *Transactions of the Institute of Measurement and Control*. 33, pp. 880-97, 2011.
- [LWT03] W. Hu L. Wang and T. Tan, Recent developments in human motion analysis. *Pattern Recognition* 36, pp. 585 - 601, 2003.
- [LZ99a] C. Loop and Z. Zhang Computing Rectifying Homographies for Stereo Vision. *In the Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, Colorado, USA, pp. 125-131, 1999.
- [LZ99b] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. *Technical report, Microsoft Research*, 1999.
- [ME95] C. Meredith and J. Edworthy, Are there too many alarms in the intensive care unit? An overview of the problems, *Journal of Advanced Nursing*, vol. 21, pp. 15-20, 1995.
- [Mer06] D. Mercier. Fusion d'informations pour la reconnaissance automatique d'adresses postales dans le cadre de la théorie des fonctions de croyance. *PhD thesis, Université de Technologie de Compiègne*, France, 2006.
- [MFB02] M. Melles, A. Freudenthal, C.A.H.M. Bouwman, Introducing Intelligent Systems into the Intensive Care Unit: a Human- Centred Approach, *21st European Conference on Human Decision Making and Control*, Scotland, pp. 110-117, 2002.
- [MFS03] C Micheloni, G.L Foresti and L. Snidaro, A co-operative multicamera system for video-surveillance of parking lots. *Intelligent Distributed Surveillance Systems Symp. by the IEE*, London, pp. 21-24, 2003.
- [MH86] R. Mohr and T. C. Henderson. Arc and Path Consistency Revisited. *Research Note, Artificial Intelligence*, pp. 225-233, vol28, 1986.
- [MLC+10] M. Monperrus, B. Long, J. Champeau, B. Hoeltzener, G. Marchalot, and J.M. Jézéquel, Model-Driven Architecture of a Maritime Surveillance System Simulator. *Systems Engineering*. Vol 13, pp. 290-7, 2010.
- [Mot06] C. Motamed, Motion detection and tracking using belief indicators for an automatic visual-surveillance system, *Image Vision Comput.* pp. 1192-1201, 2006.

- [Mot07] C. Motamed. Contribution à la conception de systèmes d'interprétation de séquences d'images, *Habilitation à Diriger les Recherches, Université du Littoral Côte d'Opale*, 2007.
- [Mot07] C. Motamed, Contribution à la conception de systèmes d'interprétation de séquences d'images, *Habilitation à dirigé les recherches, Université du Littoral de la Côte d'Opale*, 2007.
- [MQD08] D. Mercier, B. Quost, and T. Denoeux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *In Information Fusion Journal*, pp. 246-258, 2008.
- [Mur02] K. P. Murphy. Hidden semi-Markov models (HSMMs), *unpublished notes*, 2002.
- [MVB97] C. Motamed, P.Vannoorenberghe, J-M. Blosseville et J-G. Postaire La mesure du mouvement dans les séquences d'images de scènes urbaines, *4ème colloque scientifique international Transport et Pollution de l'air, Avignon, France*, pp. 260-268, 1997.
- [Nag83] H. H. Nagel. Overview on image sequence analysis. *Proc. Image Sequence Processing and Dynamic Scene Analysis*, pp. 2-39, 1983.
- [NCI02] K. C. Ng, M. Trivedi and H. Ishiguro, Generalized multiple baseline stereo and direct virtual view synthesis using range-space search, match, and render, *International Journal of Computer Vision*, pp. 131-147, 2002.
- [NCM04] H. Nait-Charif et S. J. McKenna. Activity summarization and fall detection in a supportive home environment. *Proc. of the 17th International Conference on Pattern Recognition*, pp. 323-326, 2004.
- [NGD04] D. Doermann N. Ghanem, D. DeMenthon et L. Davis, Representation and Recognition of Events in Surveillance Video Using Petri Nets, *In Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, Washington, D.C., USA, pp. 112 , 2004.
- [NGK+12] S. Nicholas, A. R. Giannareas, V. Kern, A.V. Trevino, F-M. Margaret and A. L. King, I. Lee, Smart alarms: multivariate medical alarm integration for post CABG surgery patients, *IHI ACM, Chicago, USA*, pp. 533-542, 2012.
- [NHB04] R. Nevatia, J. Hobbs, and B. Bolles. An ontology for video event representation. *In IEEE Proc. on Computer Vision and Pattern Recognition*, volume 7, pp. 119, 2004.
- [Nwa98] C. Nwagboso, User focused surveillance systems integration for intelligent transport systems, *in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): Advanced Video-based Surveillance Systems (Kluwer Academic Publishers, Boston, Chapter 1.1, pp. 8-12, 1998.*
- [OHG02] N. Oliver, E. Horvitz, and A. Garg. Layered Representations for Human Activity Recognition, *In Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, Washington, USA, pp. 830 - 856, 2002.
- [ORP00] M. Oliver, B. Rosario, and A. Pentland, A Bayesian computer vision system for modeling human interactions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 831-843, 2000.
- [OSJM07] T. Oliver, B. Schmidt, Y. Jakop, and D.L. Maskell. High speed biological sequence analysis with hidden Markov models on reconfigurable

- platforms. *IEEE Trans. on Information Theory in Biomedicine*, USA, pp. 740-746, 2007.
- [OYI92] J. Ohya J. Yamato and K. Ishii. Recognizing Human Action in Time-Sequential Images using Hidden Markov Model, *In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'92)*, Urbana Champaign, Illinois, pp. 379-385, 1992.
- [PC01] J. Piater et J. L. Crowley. Multi-Modal Tracking of Interacting Targets Using Gaussian Approximations. *Performance Evaluation for Tracking and Surveillance, PETS-2001, Kauai, Dec*, pp. 153-169, 2001.
- [Pea88] J. Pearl, Probabilistic reasoning in Intelligent Systems: Networks of plausible inference, *Morgan Kaufman*, San Mateo, Canada, pp.299-313, 1988.
- [PH01] C. Pal and M. Hu, Methodologies for constructing and training large hierarchical hidden Markov models for sequence analysis. *Computational Molecular Biology*, Pennsylvania, USA, pp. 1-2, 2001.
- [Pie07] W. Pieczynski. Multisensor triplet Markov chains and theory of evidence. *Int. Jour. Of Approximate Reasoning*, pp. 1-16, 2007.
- [Por04] F. Porikli. Trajectory distance metric using hidden Markov model based representation, *PETS Workshop, Prague*, pp, 521-524, 2004.
- [PSR98] A. Pozzobon, G. Sciutto and V. Recagno, Security in ports: the user requirements for surveillance system, *in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): Advanced Video-based Surveillance Systems (Kluwer Academic Publishers, Boston*, pp. 8-12, 1998.
- [PSV03] L. Ping, B. Lo, J. Sun, and S. A. Velastin, Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems, *Acta Automatica Sinica*, pp. 393-407, 2003.
- [PT98] M. Pellegrini, and P. Tonani, Highway traffic monitoring, *in Regazzoni, C.S., Fabri, G., and Vernazza, G. (Eds.): Advanced Video-based Surveillance Systems (Kluwer Academic Publishers)*, Boston, pp.8-12, 1998.
- [Ram07] E. Ramasso. Reconnaissance de séquences d'états par le Modèle des Croyances Transférables. Application à l'analyse de vidéos d'athlétisme *PhD thesis, Université Joseph Fourier de Grenoble 1, France*, 2007.
- [Ram09] E. Ramasso, Contribution of belief functions to Hidden Markov Models. *In IEEE Workshop on Machine Learning and Signal Processing*, pp.1-6, Grenoble, France, 2009.
- [RCC+13] P. Robert, E. Castelli, P.C Chung, T. Chiroux, C.F. Crispim-Junior, P. Mallea and F. Bremond, SWEET HOME ICT technologies for the assessment of elderly subjects, *IRBM BioMedical Engineering and Research*, Ref. No.: IRBM-D-13-00003, 2013.
- [RD00] Ronetti, N., and Dambra, C.: Railway station surveillance: the Italian case, *in Foresti, G.L., Mahonen, P., and Regazzoni, C.S. (Eds.): Multimedia Video Based Surveillance Systems, Kluwer Academic Publishers, Boston*, pp. 13-20, 2000.
- [RDS93] B.S.Y. Rao, H.F. Durrant-Whyte, J.A. Sheen, A fully decentralized multi-sensor system for tracking and surveillance, *The International Journal of Robotics Research*, Vol. 12, N°1, 1993.

- [RGR03] M. Reyes-Gomez, B. Raj, P. Daniel and W. Ellis., Multi-channel source separation by factorial HMMs, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY, USA, pp. 13-16, 2003.
- [RJM85] M.J. Russel, and R.K Moore, Explicit Modeling of State Occupancy in Hidden Markov Models for Speech Recognition , *ICASSP*, Tampa, Florida, USA, pp. 5-8, 1985.
- [RJM95] M. J. Russel, and R.K. Moore, Explicit Modeling of State Occupancy in Hidden Markov Models for Speech Recognition, *ICASSP*, Detroit, Michigan, USA, pp. 5-8.,1995.
- [RMSA+11] C. Rougier, J. Meunier, A., St-Arnaud, J. Rousseau, Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. *IEEE Transactions on Circuits and Systems for Video Technology*. 21, pp. 611-22, 2011.
- [Ros98] P. Rosin. Thresholding for change detection. *Proc. Int. Conf. Computer Vision*, Washington, DC, USA, pp. 274-279, 1998.
- [RPR+07] E. Ramasso, C. Panagiotakis, M. Rombaut, and D. Pellerin. Human action recognition in videos based on the transferable belief model - application to athletics jumps. *Pattern Analysis and Applications Journal*, Vol 11, N° 1, pp. 1-19, 2007.
- [RRL89] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. of the IEEE*, Vol 77 , pp. 257-285, Chicago, USA, 1989.
- [RRP07] E. Ramasso, M. Rombaut, and D. Pellerin, Forward-backward-Viterbi procedures in TBM for state sequence analysis using belief functions, *in Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Hammamet, Tunisia, pp. 405-417, 2007.
- [RS06] B. Ristic and Ph. Smets. The TBM global distance measure for the association of uncertain combat ID declarations. *Information Fusion*, vol 7, pp. 276-284, 2006.
- [RS98] R. Rosales et S. Sclaro, Improved tracking of multiple humans with trajectory prediction and occlusion modelling. *In IEEE CVPR Workshop on the Interpretation of Visual Motion*, Santa Barbara, Canada, pp. 1-8, 1998.
- [RT00] N. Rota and M. Thonnat. Activity Recognition from Video Sequences using Declarative Models, *14th European Conference on Artificial Intelligence (ECAI), Proceeding ECAI'00*, - W. Horn (ed.) IOS Press, Amsterdam, Berlin, Allemagne, pp. 20-25, 2000.
- [SA04] K. Sato and J. Aggarwal. Temporal spatio-velocity transform and its application to tracking and interaction. *Computer Vision and Image Understanding*, 96(2), pp. 100-128, 2004.
- [SAA12] A. Sant'Anna. A Symbolic Approach to Human Motion Analysis Using Inertial Sensors : Framework and Gait Analysis Study, *PhD thesis*, Halmstad University, Sweden, 2012.

- [San91] S. Sandri, La combinaison d'informations incertaines et ses aspects algorithmiques, *Ph.D. thesis, Université Paul Sabatier, Toulouse III, France, 1991.*
- [SAS04] A-S. Silvent, Extraction de connaissances pour la construction de scénarios médicaux, *PhD thesis, Université Joseph Fourier de Grenoble 1, France, 2004.*
- [Sch04] J. Schubert. Clustering belief functions based on attracting and conflicting metalevel evidence using Potts spin mean field theory. *Information Fusion*, pp. 309–318, 2004.
- [SCW10] P.-C., Su and C.-Y. Wu, A Joint Watermarking and ROI Coding Scheme for Annotating Traffic Surveillance Videos. *EURASIP Journal on Advances in Signal Processing*, pp. 1-14, 2010.
- [SDG05] A-S. Silvent, M. Dojat, and C. Garbay, Multi-level temporal abstraction for medical scenario construction. *International Journal Adaptive Control Signal Process*, pp. 377-394, 2005.
- [SGM07] M. Stacey, and C. McGregor, Temporal abstraction in intelligent clinical data analysis: A survey, *Artificial Intelligence in Medicine 39*, 1–24, Elsevier, 2007.
- [Sha76] G. Shafer, A Mathematical Theory of Evidence. *Princeton University Press, Princeton, New Jersey, USA, 1976.*
- [Sha81] G. Shafer, Constructive probability. *Synthese*, 48:1-60, 1981.
- [SK94] P. Smets and R. Kennes, The transferable belief model, *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.
- [SL93] R.L. Streit, T.E.Luginbuhl, A probabilistic multi-hypothesis tracking algorithm without enumeration and pruning, in *Proc. of the 6th Joint Service Data Fusion Symposium*, Laurel, Maryland, pp. 1015-1024., June 1993.
- [SL94] R.L. Streit, T.E.Luginbuhl, Maximum likelihood for probabilistic multi hypothesis tracking, *SPIE International Symposium, Orlando, USA*, pp. 394-405, April 1994.
- [SLT11] Z.L. Szpak, and J.R. Tapamo, Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert Systems with Applications.*, pp. 6669–80, 2011.
- [SMD+88] P. Smets, E. Mamdani, D. Dubois, and H. Prade, Non-Standard Logics for Automated Reasoning, *chapter Belief Functions, Academic Press*, pp. 253-286., 1988.
- [Sme02] P. Smets. The application of the matrix calculus to belief functions. *Int. Jour. of Approximate Reasoning*, 31(1-2), pp. 1–30, 2002.
- [Sme05] P. Smets. Belief functions on real numbers. *Int. Jour. of Approximate Reasoning*, vol 40, pp. 181–223, 2005.
- [Sme93] P. Smets. Beliefs functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *Int. Jour. of Approximate Reasoning*, pp. 1–35, 1993.
- [Sme98] P. Smets. The application of the Transferable Belief Model to diagnostic problems. *Int. Jour. of Intelligent Systems*, pp. 127–158, 1998.

- [SRZ11] L. Serir, E. Ramasso and N. Zerhouni, Time-Sliced Temporal Evidential Networks: the case of Evidential HMM with application to dynamical system analysis, *IEEE International Conference on Prognostics and Health Management, PHM'11*. Denver -Colorado : USA, pp. 1-10, June 2011.
- [SS03] K. Shafique and M. Shah. A non-iterative greedy algorithm for multi-frame point correspondence. *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 1, pp. 1-15, january 2005.
- [SSG+04a] L. Sebastian, V. Sventha, W. Geoff and B. Hung H., Duration abnormality detection in sequences of human activity, *Research report/technical paper, Curtin University of Technology, Perth, W. A.*, pp.1-5, 2004.
- [SSG+04b] L. Sebastian, V. Svetha W. Geoff A. W. and B. Hung H., Explicit state duration HMM for abnormality detection in sequences of human activity, *PRICAI*, Auckland, New Zealand, pp. 983-984, 2004.
- [SZP+11] A. Sumalee, R.X. Zhong, T.L. Pan and W.Y. Szeto, Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B*. 45, pp. 507-5033, 2011.
- [TBM+04] M. Thonnat, F. Bremond, N. Maillot and Van-Th. Vu. Ontologies For Video Events, *Technical Report No 5189, INRIA Sophia-Antipolis, France*, 2004.
- [Tes03] C. Tessier, Towards a commonsense estimator for activity tracking, *In AAAI Spring Symposium*, pp: 111-119, Palo Alto, CA, 2003.
- [TFS+10] M. Takahashi, M. Fujii, M., Shibata, and S.i. Satoh, Robust Recognition of Specific Human Behaviors in Crowded Surveillance Video Sequences. *EURASIP Journal on Advances in Signal Processing*, pp.1-14, 2010.
- [TH05] Y.L. Tian and A. Hampapur. Robust salient motion detection with comple background for real-time video surveillance. *Workshop on Motion and Video Computing*, Breckenridge, CO, USA, pp.30-35, 2005.
- [TR99] M. Thonnat and N. A. Rota. Image Understanding for Visual Surveillance Applications, *In International Workshop on Cooperative Distributed Vision*, pp. 51-82, Kyoto, Japan, 1999.
- [TRB98] T. Tan, P. Remagnino and K. Baker. Agent orientated annotation in model based visual surveillance, *In Proceedings of the International Conference on Computer Vision*, pp.857-862, Bombay, India, 1998.
- [TRC+06] A. Thille, P. Rodriguez, B. Cabello, F. Lellouche et L. Brochard : Patient-ventilator asynchrony during assisted mechanical ventilation. *Intensive Care Medecine*, pp.1515-1522, 2006.
- [VBT03] V.T. Vu, F. Bremond, and M. Thonnat, Automatic Video Interpretation: A Novel Algorithm for Temporal Scenario Recognition, *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*, pp. 9-15, Acapulco, Mexico, 2003.
- [VT04] V.T. Vu, Temporel scenario for automatic video interpretation, *PhD thesis, INRIA, Sophia-Antipolis, France*, 2004.

- [VTB02] V-T. Vu, F. Bremond and M. Thonnat, Temporal Constraints for Video Interpretation, *The 15-th European Conference on Artificial Intelligence (ECAI'2002)*, France, 2002.
- [VVA04] M. Valera, S. A. Velastin, Intelligent distributed surveillance systems: A review, *IEEE Proceedings*, pp. 193–204, 2004.
- [VVA05] M. Valera and S.A. Velastin, Intelligent distributed surveillance systems: a review, *IEEE Proceedings on Intelligent Distributed Surveillance Systems*, pp. 192-204, 2005.
- [Wan12] X. Wang, Intelligent multi-camera video surveillance: A review, *Pattern Recognition Letters*, Elsevier, pp. 3-19, 20 July 2012.
- [WB01] A. Wilson and A. Bobick. Hidden Markov models for modeling and recognizing gesture under variation. *Int. Jour. of Pattern Recognition and Artificial Intelligence*, pp.123–160, 2001.
- [WES95] M.E. Weber, and M.L Stone, Low altitude wind shear detection using airport surveillance radars, *IEEE Aerosp. Electron. Syst. Mag.*, 10, (6), pp. 3–9,1995.
- [Wix00] L. Wixson. Detecting salient motion by accumulating directionally consistent flow. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8) :774-780, 2000.
- [WPV05] G. West P. Peursum and S. Venkatesh. Combining image regions and human activity for indirect object recognition in indoor wide-angle views, *In IEEE International Conference on Computer Vision*, Vol. 1, pp. 82-89, 2005.
- [WTK+11] W.R. Webster, E. Töpfer, F.R. Klauser and C.D. Raab, Revisiting the surveillance camera revolution: Issues of governance and public policy. *Introduction to part one of the Special issues. Information Polity*. 16, pp: 297–301, 2011.
- [XLO04] M. Xu, L. Lowey, and J. Orwell, Architecture and algorithms for tracking football players with multiple cameras. *Proc. IEE Workshop on Intelligent Distributed Surveillance Systems*, London, England, pp. 51–56, 2004.
- [Yag87] R.R. Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, pp :93–137, 1987.
- [YCB11] M. C. Yuksek and B. Barshan. Human activity classification with miniature inertial and magnetic sensors. *In 19th European Signal Processing Conference*, pp: 956–960, Barcelona, Espagne, 2011.
- [YJS06] A. Yilmaz, O. Javed, and M. Shah,. Object tracking: A survey. *ACM Comput. Surv.* 38, 4, Article 13, pp: 1-45, Dec. 2006.
- [YSM01] B. Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. The marginal case. *Int. Jour. of Approximate Reasoning*, pp.47–70, 2001.
- [YSM02a] B. Yaghlane, P. Smets, and K. Mellouli. Belief function independence: II. The conditional case. *Int. Jour. of Approximate Reasoning*, pp :31–75, October 2002.
- [YSM02b] B. Yaghlane, P. Smets, and K. Mellouli. Independence concept for belief functions. *In Technologies for constructing intelligent systems: Tools. Physica-Verlag*, pp. 31-75, 2002.

- [YZY⁺11] G. Yuan, X. Zhang, Q. Yao, and K. Wang, Hierarchical and Modular Surveillance Systems. *in ITS IEEE Intelligent Systems*. 26, pp. 10-5, 2011.
- [ZBT10] N. Zouba, F. Bremond and M. Thonnat, An Activity Monitoring System for Real Elderly at Home: Validation Study. *In the 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS*, Boston, pp.1-8, 2010.
- [Zha00] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334, 2000.
- [Zia10] A. Ziani. Interprétation en temps réel de séquence vidéo par exploitation des modèles graphiques probabilistes, *Phd thesis, Université du Littoral Côte d'Opale*, France, 2010.
- [ZM11] A. Ziani, and C. Motamed, Automatic Scenario Recognition for Visual-Surveillance by Combining Probabilistic Graphical Approaches, *Book Video Surveillance ISBN 978-953-307-436-8 Edited by: Weiyao Lin, Publisher: InTech*, pp. 322-334, 2011.
- [ZMN08] A. Ziani, C. Motamed, and J.C. Noyer, Temporal reasoning for scenario recognition in video-surveillance using Bayesian networks, *IET Computer Vision*, vol. 2, no 2, pp. 99-107, 2008.
- [ZSQ⁺11] J. Zhang, G. Song, G. Qiao, T. Meng and H. Sun, An Indoor Security System with a Jumping Robot as the Surveillance Terminal. *IEEE Transactions on Consumer Electronics*, pp. 1774-1781, 2011.

Résumé

Les travaux de recherche développés dans cette thèse concernent la mise en œuvre d'un système de vidéo surveillance intelligente en milieu hospitalier. Dans le contexte d'une application en unité de soins intensifs médicale, nous introduisons la notion originale de Boite Noire Médicale et nous proposons un nouveau système de monitoring visuel de Détection Automatique de Situations à risque et d'Alerte (DASA) basé sur un système de vidéosurveillance multi caméra intelligent. L'objectif étant d'interpréter les flux d'informations visuelles et de détecter en temps réel les situations à risque afin de prévenir l'équipe médicale et ensuite archiver les événements dans une base de donnée vidéo qui représente la Boite Noire Médicale. Le système d'interprétation est basé sur des algorithmes de reconnaissance de scénarios qui exploitent les Modèles de Markov Cachés (MMCs). Une extension du modèle MMC standard est proposée afin de gérer la structure hiérarchique interne des scénarios et de contrôler la durée de chaque état du modèle markovien. La contribution majeure de ce travail repose sur l'intégration d'un raisonnement de type évidentiel, pour gérer la décision de reconnaissance en tenant compte des imperfections des informations disponibles. Les techniques de reconnaissance de scénarios proposées ont été testées et évaluées sur une base de séquences vidéo médicales et comparées aux modèles de Markov cachés probabilistes classiques.

Mots-clés : *Interprétation de séquences d'images, Reconnaissance de scénarios, Modèles de Markov cachés, Modèle des Croyances Transférables, Raisonnement temporel.*