



HAL
open science

Vidéosurveillance d'une salle de traitement en radiothérapie

Miguel Alfredo Portela Sotelo

► **To cite this version:**

Miguel Alfredo Portela Sotelo. Vidéosurveillance d'une salle de traitement en radiothérapie. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Claude Bernard - Lyon I, 2011. Français. NNT : 2011LYO10260 . tel-03510286

HAL Id: tel-03510286

<https://theses.hal.science/tel-03510286>

Submitted on 4 Jan 2022

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.

THÈSE

pour obtenir le grade de
DOCTEUR de U.C.B.L.
Spécialité : **Informatique**

présentée et soutenue publiquement
par

Miguel Alfredo PORTELA SOTELO

le 16 décembre 2011

VIDÉOSURVEILLANCE D'UNE SALLE DE TRAITEMENT EN RADIOTHÉRAPIE

Jury

Kadi BOUATOUCH	Rapporteur	Professeur Université de Rennes I
Franck MARZANI	Rapporteur	Professeur Université de Bourgogne – Dijon
Jocelyne TROCCAZ	Examineur	Directeur de recherche CNRS Université Joseph Fourier – Grenoble
Pascal POMMIER	Examineur	Radiothérapeute (Docteur en médecine) Centre Léon Bérard – Lyon
Hanna KAFROUNI	Examineur	P.D.G. (Docteur – HDR) DOSIsoft S.A. – Cachan
Chantal GINESTET	Invitée	Physicienne Médical Centre Léon Bérard – Lyon
Jean-Michel MOREAU	Directeur de thèse	Professeur Université Claude Bernard Lyon 1
Élodie DESSERÉE	Co-directeur de thèse	Maître de conférences Université Claude Bernard Lyon 1

Thèse préparée au **Laboratoire d'InfoRmatique en Images et Système d'information**
dans le cadre de l'École Doctorale **Math-Info**
et d'une bourse CIFRE avec l'entreprise **DOSIsoft S.A.**

À mes cousines
Monica et Louise,
parties trop tôt ;
pour que la lutte contre
le cancer puisse permettre à
de si merveilleuses personnes
de rester avec nous.

Résumé

Les différents acteurs impliqués dans le traitement du cancer par radiothérapie (oncologues, radiophysiciens, manipulateurs, etc.) ont exprimé un réel besoin d'avoir une vue globale du système de traitement pour pouvoir suivre le patient ainsi que la totalité des équipements de la salle - mobiles (voire robotisés) ou non - afin d'améliorer la sécurité des patients. Cette vue globale permettrait de réduire le nombre d'interventions humaines, les erreurs d'ordre physique (mouvements involontaires, collisions, etc.) et les erreurs dans la configuration globale par rapport au patient traité.

Pour répondre à ce besoin, nous présentons dans cette thèse un module de vidéosurveillance exploitant un modèle numérique des éléments de la scène (patient, couche, bras d'irradiation, etc.) afin de suivre le mouvement de ces éléments dans le but d'éviter tout événement indésirable susceptible de mettre en danger le patient, le personnel voire les équipements. Surveiller la salle de traitement consiste à vérifier, à tout instant, que sa configuration courante est conforme à ce qui a été planifié en amont (dans le Système de Planification de Traitement [SPT]), ainsi qu'aux normes liées à la radiothérapie, et le cas échéant, à informer le personnel médical pour qu'il prenne les mesures nécessaires.

Nous avons développé un modèle générique, qui permet la création de modèles numériques à partir des données contenues dans le dossier du patient (issues du SPT), enrichies d'autres informations visuelles, géométriques et « sémantiques ». L'ensemble des modèles numériques permet de créer un environnement virtuel équivalent à la scène vue par les caméras vidéo. La localisation et l'orientation des éléments est calculée par recalage 3D, choix justifié par les caractéristiques propres aux salles de traitement par radiothérapie (milieu contrôlé, dynamique faible, spécification de construction très contrainte : salles très semblables, etc.). Les informations (notamment géométriques, physiques et mécaniques) contenues dans le modèle permettent de réduire la complexité globale des algorithmes de contrôle.

La méthode de suivi de mouvement implémentée exploite les caractéristiques et les contraintes des éléments incluses dans leur modèle numérique, et plus précisément les « descripteurs », qui correspondent à la modélisation 3D des primitives permettant de suivre l'élément dans le flux vidéo. Nous présentons également une fonction de dissimilarité, que nous avons définie, résultant de la fusion de deux fonctions de dissimilarité bien connues : distance de chanfrein et notion de non recouvrement. Un formalisme flou permet d'améliorer la robustesse de la méthode face au bruit

généralisé par le système d'acquisition et le traitement des données. Lorsque la séance de traitement se déroule, le module évalue la configuration courante (positions acquises des éléments à l'aide du module de suivi et des paramètres globaux de la scène), pour interagir avec le personnel médical par rapport à ce qui a été défini en prétraitement.

Nous présentons des résultats issus de tests réalisés dans une salle de radiothérapie mais en l'absence d'irradiation et de patient (un fantôme positionné sur la couche s'est substitué au patient, ou bien l'un de nos collaborateurs a joué le rôle du patient). Ces résultats démontrent la faisabilité et l'efficacité du module de surveillance.

Mots-clefs Réalité augmentée et virtuelle, modélisation, vision par ordinateur, suivi de mouvement, logique floue, analyse vidéo, reconnaissance de scénarios, assurance qualité, gestion de risques, interactivité, radiothérapie.

Abstract

The various professionals (oncologists, medical physicists, radiation therapist, etc.) involved in the treatment of cancer by radiation therapy express more and more acutely a real need to have a global view of the treatment system in order to monitor the patient and all the equipment in the room — be it mobile (or even robotized) or not — to improve patient safety. This global view would help reduce the number of human interventions, errors in physical interaction (involuntary movements, collisions, etc.), and errors in the global configuration for the patient.

To answer this need, this PhD manuscript presents a videosurveillance module using a numerical model of the elements in the scene (patient, couch, irradiation arm, etc.) to monitor the movement of these elements in order to avoid any event that could put the patient, staff or equipment in danger. Monitoring the treatment room means checking, at any moment, that its current configuration is consistent with what was planned upstream (in the Treatment Planning System [TPS]) and with the standards of radiotherapy, and if need be to, to inform the medical staff, so that correctiong actions may be undertaken.

We have developed a generic model which allows the creation of numerical models from the data contained in the patient record (in the TPS), enriched with other visual, geometric and “semantic” information. All the numerical models used to create a virtual environment correspond to the scene as seen by video cameras. Location and orientation of the elements are calculated by 3D registration, a choice justified by the characteristics of radiation treatment rooms (controlled environment, low dynamics, highly constrained construction specifications, very similar treatment rooms, etc.). The (geometric, physical and mechanical) information items in the model allow to reduce the overall complexity of the algorithms.

The method for motion tracking we have implemented exploits the characteristics and constraints of the elements included in the numerical model, specifically the “descriptors” which correspond to the 3D modeling of primitives used to track the element in the video stream. We also present a dissimilarity function that we have defined, resulting from the merging of two well-known dissimilarity functions, respectively based on the chamfer distance and the notion of non-overlap. A fuzzy formalism improves the robustness of the method against the noise generated by acquisition and processing. When the treatment session takes place, the module evaluates the current configuration (positions acquired for elements, using the tracking module and global parameters of the scene), to interact with medical staff in relation to what has been defined in pretreatment.

We present results of tests performed in a radiotherapy room but in the absence of irradiation and patient (the patient was either replaced by a phantom positioned on the couch or by one researcher playing his/her role). These results demonstrate the feasibility and efficiency of the surveillance module.

Keywords Augmented and virtual reality, modeling, computer vision, motion tracking, fuzzy logic, video analysis, scenario recognition, quality assurance, risk management, interactivity, radiotherapy.

Resumen

Los profesionales involucrados en el tratamiento contra el cáncer con radioterapia (oncólogos, radiofísicos hospitalarios, tecnólogos médicos, etc.) expresaron la necesidad de tener una vista global del sistema de tratamiento con el fin de supervisar el paciente y todos los equipos presentes en la habitación - móvil (o robot) o no - para mejorar la seguridad del paciente. Esta vista global reduciría la cantidad de intervenciones humanas, los errores físicos (movimientos involuntarios, colisiones, etc.) y los errores en la configuración global con respecto al paciente tratado.

Para satisfacer esta necesidad, en esta tesis presentamos un módulo de videovigilancia que explota un modelo digital de los elementos de la escena (el paciente, la camilla, la máquina de radiación, etc.) para vigilar el movimiento de estos elementos con el fin de evitar cualquier evento adverso que podría poner en peligro al paciente, el personal o los equipos. La supervisión de la sala de tratamiento implica la verificación, en cualquier momento, la consistencia entre la configuración actual de la sala con respecto a lo que estaba previsto antes (en el Sistema de Planificación del Tratamiento [SPT]) y con respecto a las normas relacionadas con la radioterapia, y en el caso contrario, informar al personal médico para que el tome las medidas necesarias.

Hemos desarrollado un modelo genérico, que permite la creación de modelos digitales de los datos contenidos en el registro del paciente (obtenidas con el SPT), enriquecidas con otro tipo de información visual, geométrica y « semántica ». Los modelos digitales de todos los elementos son utilizados para crear un entorno virtual equivalente a la escena como es visto por las cámaras de vídeo. La ubicación y orientación de los elementos se calcula con la técnica de registro 3D, elección justificada por las características de las salas de tratamiento con radioterapia (entorno controlado, dinámica baja, especificación de la construcción con muchas restricciones : salas muy similares, etc.). Las informaciones (principalmente geométricas, físicas y mecánicas) contenidas en el modelo permiten la reducción de la complejidad general de los algoritmos utilizados.

El método de seguimiento del movimiento implementado explota las características y las limitaciones de los elementos incluidas en sus modelos digitales, y plus exactamente los « descriptores », que corresponden al modelado 3D de las primitivas que permiten controlar el elemento de la secuencia de vídeo. También se presenta una función de disimilitud, que nosotros hemos definido, resultante de la fusión de dos funciones de disimilitud muy bien conocidas : distancia de chaflán y la noción de no-recubrimiento. Un formalismo difuso, basado en la teoría de conjuntos difusos, mejora la robustez del método contra el ruido generado por la adquisición y el procesamiento de datos. Cuando la sesión de tratamiento se lleva a cabo, el módulo evalúa la configuración actual (las posiciones adquiridas de los elementos mediante el módulo de seguimiento y los parámetros globales de la escena), para interactuar con el personal médico en relación con lo que se ha definido antes del tratamiento.

Presentamos los resultados de las pruebas realizadas en una sala de radioterapia, pero en ausencia de la radiación y del paciente (un objeto reemplazado al verdadero paciente ha sido colocado sobre la camilla, o una persona de nuestro equipo ha jugado el papel del paciente). Estos resultados demuestran la viabilidad y la eficacia del módulo de seguridad.

Palabras claves La realidad aumentada y virtual, modelado digital, la visión artificial, seguimiento del movimiento, lógica difusa, análisis de vídeo, reconocimiento de escenario, aseguramiento de la calidad, gestión de riesgos, interactividad, radioterapia.

REMERCIEMENTS

Je voudrais d'abord remercier mes encadrants, M. Jean-Michel MOREAU et Mme Élodie DESSERRÉE, pour avoir dirigé cette thèse, pour leur forte implication tant scientifique qu'humaine, pour leurs conseils avisés et pour leur confiance. J'ai énormément apprécié de travailler avec eux et de bénéficier de leur expérience que j'ai mise à profit.

Je remercie M. Hanna KAFROUNI, Président Directeur Général de la société DOSIsoft, pour l'intérêt tout particulier qu'il a porté à ces travaux, pour son expertise et pour avoir rendu possible cette thèse grâce à son soutien financier. Je remercie également l'ensemble des employés de DOSIsoft, notamment MM. Baptiste GERMAIN, Jean-Christophe DIAZ et Christopher TARENTO, pour leur accueil et pour les échanges scientifiques que nous avons pu avoir au cours de ces trois années de thèse. Je remercie également M. Antony PEREIRA, stagiaire de la société DOSIsoft pour son aide dans le cadre de ces travaux..

J'aimerais aussi remercier Mmes Chantal GINESTET et Pauline DUPUIS, radiophysiciennes au Centre Léon Berard (CLB), pour leur soutien, leur immense apport et savoir-faire dans le domaine de la radiothérapie, et pour m'avoir permis de tester et de valider mes travaux dans le locaux du CLB. Je tiens à remercier tout particulièrement Pauline DUPUIS, sans qui l'intégration de la partie médicale n'aurait pas pu exister, pour sa forte implication dans ces travaux et sa grande disponibilité. Je remercie également le corps médical du service de radiothérapie au CLB, pour son accueil et son aide.

Un grand merci à tous les membres du jury pour avoir accepté d'évaluer cette thèse, et plus particulièrement aux Professeurs Franck MARZANI et Kadi BOUATOUCH pour leur investissement en tant que rapporteurs.

Merci à tous les membres du projet ETOILE et du laboratoire LIRIS, plus spécialement les membres de l'équipe SAARA, pour leur aide et leur soutien à chaque fois que j'en ai eu besoin, et pour l'accueil très chaleureux. Je tiens à remercier tout particulièrement Alexandre, Brice et Ouissem, qui ont si souvent montré leur intérêt à mon travail, compris mes besoins et fait des suggestions sans qu'il soit nécessaire de leur demander. Merci à Aurélien, Jacques, Jeremy, Maxime et Adrien d'avoir rendu très agréable le temps passé au laboratoire, et en dehors aussi.

Merci à mon épouse, Heather, pour son soutien inconditionnel qui m'a toujours motivé à ne pas baisser les bras, et pour la joie qu'elle m'apporte au quotidien. Un grand merci à mes parents, ma sœur et toute ma famille, ainsi qu'à la famille de mon épouse, David, Agnès, et à mes amis proches, Peter, Daniel, Rebeca et Stephanie, pour leur aide, leur soutien et leur amour.

Enfin, merci à toute la communauté française pour m'avoir accueilli si chaleureusement et merci à tous ceux qui ont toujours cru en moi.

Miguel Alfredo Portela Sotelo

Table des matières

Résumé	iii
Table des matières	i
1 Introduction	1
1.1 Erreurs en radiothérapie	3
1.1.1 Incidents lors d'une séance de traitement	4
1.2 Besoins	5
1.2.1 Contraintes	6
1.3 Objectifs	7
1.3.1 Problématiques à considérer	7
1.4 Contextes académique et industriel	9
2 État de l'art	11
2.1 Radiothérapie et imagerie médicale	11
2.1.1 Techniques d'imagerie	13
2.1.2 Plan de traitement	15
2.1.3 Déroulement des séances	18
2.2 Assistance au positionnement du patient	19
2.2.1 Systèmes d'immobilisation	19
2.2.2 Systèmes de contrôle lors du positionnement	20
2.2.3 Systèmes de contrôle lors de l'irradiation	24
2.3 Contrôle en radiothérapie	25
2.3.1 Assurance Qualité en radiothérapie	26
2.3.2 Gestion des risques	27
2.3.3 Contrôle du risque humain	28
2.4 Méthodes de suivi d'individus et d'objets	29
2.4.1 Représentation numérique de l'élément	29
2.4.2 Techniques d'acquisition du mouvement	30
2.4.3 Suivi du mouvement 3D de plusieurs éléments	34
2.5 Vidéosurveillance	35
2.5.1 Vidéosurveillance automatisée	35
2.5.2 Surveillance du patient dans le domaine médical	36
2.6 Réalité Augmentée	37
2.6.1 Réalité Augmentée dans le domaine médical	37

3	Représentation de l'environnement	39
3.1	Approche générale	39
3.2	Système d'acquisition	40
3.2.1	Calibrage du système	40
3.2.2	Modèle numérique du système d'acquisition	43
3.3	Modèle générique	43
3.3.1	Description de la structure	44
3.3.2	Intégration des informations dans le modèle	45
3.3.3	Intégration du modèle dans le module du contrôle	46
3.3.4	Descripteurs 3D	46
3.3.5	Opérateurs dans le modèle numérique	48
3.3.6	Initialisation du modèle personnalisé	49
3.4	Environnement de Réalité Augmentée	49
4	Calcul de la pose	51
4.1	Pose 3D	51
4.2	Recalage	53
4.2.1	Recalage 3D pour le suivi de mouvement	55
4.3	Formalisation de la solution proposée	56
4.3.1	Algorithme global de la méthode	57
4.3.2	Modèle temporel du flux vidéo	59
4.3.3	Détection de mouvements	61
4.4	Fonction de dissimilarité	62
4.4.1	Carte de distances	66
4.5	Méthodologie générale du calcul de pose	67
4.5.1	Évaluation des descripteurs pour la calcul d'une pose	68
4.5.2	Pondération de la pose	68
4.5.3	Algorithme d'optimisation	71
4.5.4	Utilisation des files de priorité	73
5	Reconnaissance d'événements et de scénarios	77
5.1	Gestion des erreurs en radiothérapie par vidéo	78
5.1.1	Utilisation de la vidéo dans une salle de traitement	79
5.1.2	Vidéosurveillance automatisée d'une salle de traitement	79
5.2	Modélisation d'une séance de traitement	81
5.3	Reconnaissance d'événements	85
5.3.1	Classification d'un événement	86
5.3.2	Analyse des événements	88
5.4	Reconnaissance de scénarios	90
5.4.1	Application à la vidéosurveillance de la salle	91
5.4.2	Modèle de scénario	92

6	Module de contrôle externe	97
6.1	Traçabilité	97
6.2	Interaction avec l'utilisateur	98
6.2.1	État global de la séance	99
6.2.2	Écrans de vidéosurveillance	100
6.2.3	Configurations	101
6.2.4	Frise chronologique	103
6.3	Interaction des sous-modules	103
6.4	Implémentation du module	104
6.5	Mise en place du module	105
6.5.1	Salle de traitement et modèles numériques des éléments	106
7	Validation et résultats	109
7.1	Modèle générique	109
7.1.1	Modélisation d'une salle de traitement	111
7.1.2	Conclusions sur le modèle générique	112
7.2	Suivi de mouvement	112
7.2.1	Résultats des tests sur des données synthétiques	112
7.2.2	Résultats des tests sur des données réelles	117
7.2.3	Conclusions sur la méthode de suivi de mouvement	121
7.3	Reconnaissance et analyse d'événements	122
7.3.1	Résultats	122
7.3.2	Conclusions sur le traitement d'événements	124
8	Conclusion générale	127
8.1	Perspectives	128
	Bibliographie	131
	Références web	153

Chapitre 1

Introduction

Le cancer, maladie caractérisée par le développement par prolifération de cellules d'une tumeur maligne, est l'une des principales cause de décès dans le monde : 7,6 millions de personnes en 2008 et une estimation de 12 millions de décès en 2030 d'après le CIRC [1]. Plusieurs types de traitements sont possibles : l'ablation chirurgicale pour extraire la tumeur, l'utilisation de la radiofréquence, de micro-ondes, de la cryothérapie, de la chimiothérapie, ou de la radiothérapie pour détruire la tumeur. Environ la moitié des malades du cancer sont traités par la radiothérapie externe, qui consiste à utiliser les rayonnements ionisants produits par un accélérateur de particules à usage médical. Cette méthode de traitement a été créée au début du 20ème siècle et, depuis, elle a constamment évolué au rythme des innovations technologiques. Ces traitements complexes demandent de plus en plus l'intégration de technologies d'autres domaines que le médical, faisant aujourd'hui cohabiter des recherches pluridisciplinaires dans un but commun : la lutte contre le cancer. L'informatique a fourni de grands apports à la radiothérapie, notamment grâce à des logiciels de gestion (patients, personnel, administratif, etc.), de planification du traitement, de traitement d'images médicales, de simulation, etc..

L'assurance qualité en radiothérapie porte localement sur différentes étapes du traitement de radiothérapie (Fig. 1.1) et globalement sur les dispositifs médicaux utilisés tout au long de la chaîne. Nos travaux se focalisent sur les étapes de positionnement du patient et de la délivrance du traitement, et exploitent les informations obtenues lors des étapes précédentes.

En radiothérapie, un progiciel assure la sécurisation de la transmission à l'accélérateur des données du traitement de chaque patient pendant toutes les étapes du traitement, et enregistre automatiquement les données de chaque séance d'irradiation. Ce type de système est appelé *Record and Verify* (en français, Enregistrement et Vérification), bien qu'en réalité il vérifie d'abord les paramètres internes des machines par rapport aux données planifiées, principalement les informations concernant les caractéristiques des faisceaux d'irradiation (énergie, forme, balistique) et la dose prescrite, afin de détecter des erreurs. À la fin de l'irradiation, le système enregistre ces paramètres dans des fichiers « log », qui pourront être consultés ultérieurement si besoin. Ce type de système est largement répandu dans les centres de radiothérapie pour réduire les risques. Ces centres utilisent également des systèmes complémentaires permettant contrôler le positionnement du patient.

Malgré tout, radiothérapeutes et physiciens médicaux expriment de plus en plus

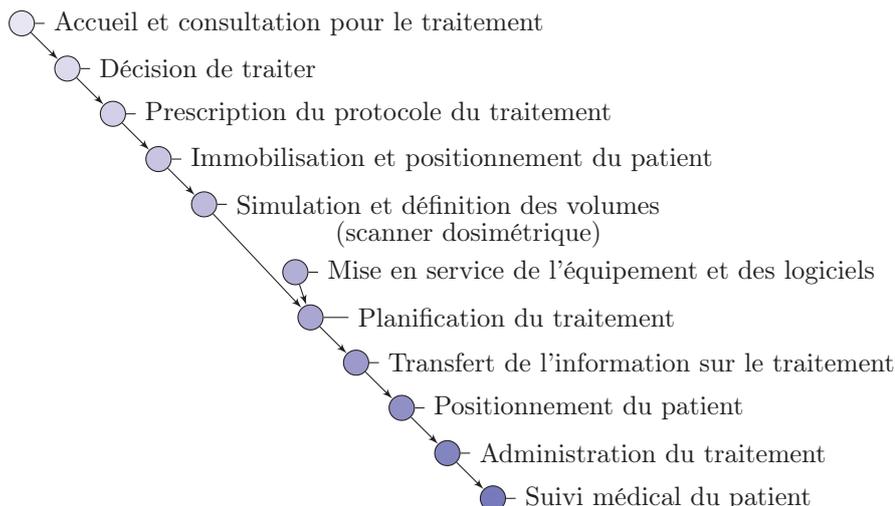


FIGURE 1.1: Étapes de la prise en charge du patient en radiothérapie [mdls09].

le besoin d'avoir une vue globale externe de la salle de traitement (patient et environnement) afin de réduire les erreurs potentielles que les systèmes de *Record and Verify* actuels et les autres systèmes existants ne sont pas capables de détecter lors d'une séance. Cette vue globale externe comprend le suivi du mouvement de tous les éléments dans la salle dans un référentiel unique. Cet outil doit permettre également de réaliser l'analyse des positions des éléments et autres paramètres externes, afin de vérifier leur conformité¹ avec ce qui a été planifié. De plus, l'analyse des informations acquises peut également répondre à des problématiques liées à la radioprotection du personnel (présence ou non du personnel à l'intérieur de la salle avant de démarrer l'irradiation?).

Nos travaux répondent à ce besoin par le développement d'un module de contrôle externe composé d'un système de caméras vidéo couplé à un système logiciel utilisant des techniques de vision par ordinateur, et plus particulièrement de vidéosurveillance. Cette approche a été jusqu'à maintenant peu prise en compte pour répondre aux problèmes d'assurance qualité en radiothérapie. L'idée générale est d'exploiter les informations déjà contenues dans le dossier médical du patient (anatomie du patient dans la zone à irradier, caractéristiques physiologiques, etc.) et de vérifier à tout instant leur conformité par rapport à ce qui est vu par le système de caméras.

Dans la suite de ce premier chapitre, nous présenterons le contexte, les objectifs et les problématiques à résoudre. Le chapitre 2 sera consacré à l'état de l'art sur les problématiques mises en évidence dans la section 1.3.1, et nous présenterons en détails les principes généraux au début du chapitre 3. Notre module de contrôle externe est composé de trois parties principales. La première partie de ce module sera présentée à la fin du chapitre 3, et elle correspond au modèle numérique qui a été développé afin d'exploiter les informations *a priori* dont on dispose sur le patient, le personnel, la salle de traitement et les objets potentiellement présents dans la salle. La deuxième partie du module, présentée au chapitre 4, correspond au suivi de mouvement de plusieurs éléments dans la scène. La troisième et dernière

1. La conformité est par définition (ISO 9000) la satisfaction d'une exigence.

partie sera présentée au chapitre 5, et elle correspond à la reconnaissance d'événements et de scénarios. Le chapitre 6 clôturera les explications sur la conception et le développement du module de contrôle externe, notamment avec des informations sur l'interaction entre les différents composants, le modèle numérique, le système de caméras et l'utilisateur. Les résultats obtenus seront présentés au chapitre 7, et des conclusions seront données au chapitre 8.

Les termes propres à la radiothérapie, utilisés dans ce chapitre, seront détaillés au début du chapitre 2. Tout au long de ce document, les citations comportant uniquement des numéros correspondent aux références web.

1.1 Erreurs en radiothérapie

En radiothérapie, la gestion des erreurs est l'une des priorités des organismes de santé et de radioprotection dans le monde. Des organismes comme l'ASTRO [2], l'ESTRO [3], l'ASN [4] et l'OMS [5] ont publié des rapports avec des conseils et des recommandations pour le patient et son entourage, et des consignes pour le personnel médical, afin de les sensibiliser à cette problématique [mdls09, Soc10, dSN11].

Actuellement, la gestion des risques en radiothérapie est faite en deux temps : d'abord par une analyse de risques *a priori* qui détermine les points de contrôle et de validation obligatoires, puis par retour d'expérience du personnel médical et médico-technique. Il est notamment mis en évidence que l'erreur est humaine, mais que l'on ne peut pas tolérer sa répétition. Sur ce principe, des systèmes ont été mis en place afin de répertorier les incidents observés lors des traitements [CCKH10, DBN⁺11]. Cependant, répertorier les incidents demande un travail supplémentaire pour le personnel et une remise en question régulière des pratiques du service. Les organismes de santé et de radioprotection ont répondu en partie à cette problématique en établissant des procédures et des modes opératoires à suivre.

Dans le cas idéal, on souhaiterait prédire une erreur potentielle et réduire, voire éliminer toutes les causes d'erreur. Nos travaux, qui vont dans cette direction, se focalisent sur la détection de paramètres pouvant produire une erreur, afin d'assister le personnel médical (en lui indiquant une source potentielle d'erreur) et de répertorier automatiquement l'incident. Le système proposé s'intéresse aux paramètres externes qui sont pris en compte visuellement par le personnel chargé de l'administration du traitement, mais non par les outils logiciels actuellement disponibles. Des exemples de ces paramètres sont la localisation des individus (pas seulement le patient) et la position relative des objets (machines notamment) présents dans la salle de traitement. En utilisant des images vidéo, nous apportons de nouvelles informations, que les systèmes actuels (équipement et logiciel) en radiothérapie ne sont pas capables de fournir.

Les erreurs sont rares en radiothérapie, mais elles peuvent avoir des effets cliniques néfastes sur le patient. Ces effets peuvent être différés dans le temps, et dans ce cas, une simulation dosimétrique est utilisée pour évaluer les conséquences potentielles liées à l'écart entre la dose prescrite et la dose délivrée ainsi que à l'écart entre le volume cible et le volume réellement irradié. En effet, il existe des valeurs de dose limites à partir desquelles on constate des symptômes sur le patient. De plus, les différents organes n'ont pas la même radiosensibilité : pour les « Organes

À Risque » (OAR) la dose de tolérance fixée est « infranchissable » (moelle épinière par exemple). L'ASN [4], en concertation avec la Société Française de Radiothérapie Oncologique (SFRO) [6], a publié en 2008 l'échelle ASN-SFRO pour la prise en compte des événements de radioprotection affectant des patients dans le cadre d'une procédure de radiothérapie [dSN10].

Entre 1976 et 2007, aux États Unis et dans les pays d'Amérique Latine, d'Europe et d'Asie, environ 1500 millions de patient ont été traités, dont 3125 ont eu des résultats (confirmés) indésirables et non compensables dans leur traitement à cause d'une erreur détectée et identifiée [SBN⁺09]. Sur la Fig. 1.2, ces erreurs sont classées par rapport aux étapes du processus du traitement présentées dans la Fig. 1.1. Ce rapport n'indique pas les erreurs détectées et corrigées, ni les erreurs non détectées, mais il permet de mettre en évidence la gravité d'une erreur n'ayant pas été traitée. En 2010 en France, l'ASN a reçu 116 rapports d'événements de niveau 1 sur l'échelle ASN-SFRO (événement avec conséquence dosimétrique mais sans conséquence clinique attendue), sur les 164 000 traitements qui ont été réalisés cette même année.

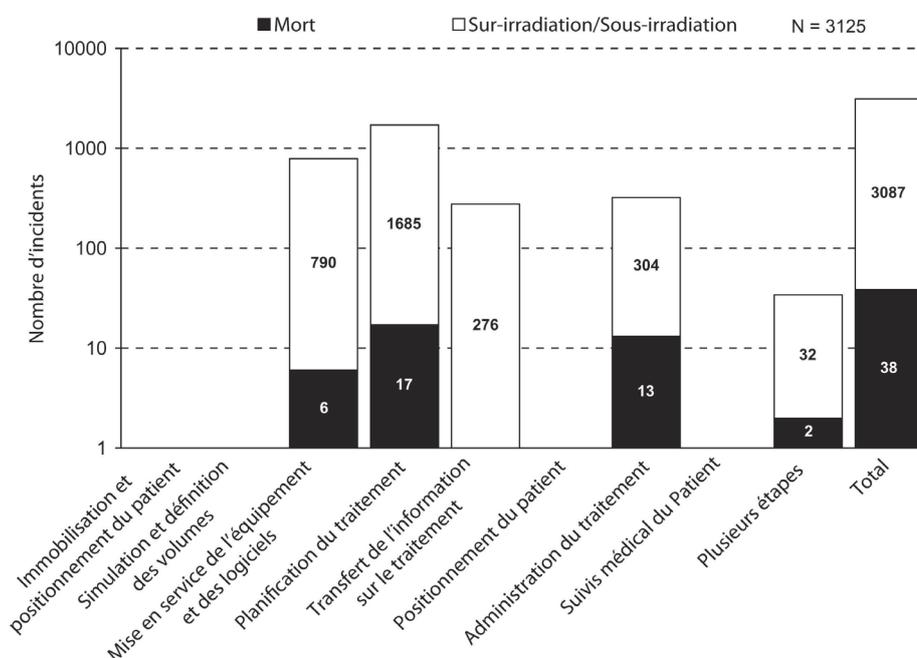


FIGURE 1.2: Incidents en radiothérapie avec des résultats indésirables et non compensables pour le patient, par étape du processus (1976–2007) [SBN⁺09].

1.1.1 Incidents lors d'une séance de traitement

Pour la planification du traitement, une séance d'acquisition d'images de tomodensitométrie (CT Scan) est réalisée. Le patient est d'abord placé sur la table du scanner pour la dosimétrie et dans une position qui sera définie comme **référence** pour toute la suite du traitement. Des moyens de contention sont utilisés afin d'immobiliser le patient et de rendre reproductible cette position de référence du patient. Ces systèmes n'éliminent cependant pas toutes les imprécisions, liées au positionnement, qui seront compensées par l'utilisation de volumes de sécurité (marges). Les images sont ensuite transférées au Système de Planification de Traitement (SPT,

ou TPS pour Treatment Planning System), sur lequel le radiothérapeute trace la tumeur à irradier et les régions saines à éviter. Ensuite, un plan de traitement est conçu par les physiciens médicaux à l'aide du SPT. Ce plan contient notamment les informations sur les caractéristiques des faisceaux et la dose prévisionnelle calculée qui sera délivrée à la tumeur, la balistique choisie permettant de préserver le plus possible les tissus sains environnants.

Une séance de traitement est constituée de deux étapes : le positionnement du patient et la délivrance du traitement (dose d'irradiation). Le positionnement consiste à placer le patient sur la table de traitement dans la même position de référence qu'il avait lors de la séance d'acquisition d'images CT. De plus, dans le référentiel de l'accélérateur cette position doit être conforme à ce qui a été défini dans le plan de traitement. Une fois cette tâche réalisée, on peut passer à l'étape de la délivrance du traitement qui consiste à irradier la tumeur comme cela a été planifié. Plusieurs facteurs entrent en jeu mais l'un des plus importants est d'épargner les tissus sains. Pour ceci, la balistique choisie en dosimétrie est généralement définie par un ensemble d'incidences (angles) d'irradiation, afin de réduire l'irradiation des tissus sains et des OAR et de concentrer au mieux la dose déposée dans la tumeur. Plusieurs faisceaux sont donc délivrés, chacun ayant ses propres paramètres et pouvant requérir l'utilisation d'un accessoire modificateur du faisceau au niveau du patient ou du collimateur. De plus, le patient doit rester immobile afin que le traitement soit délivré correctement.

Le fractionnement du traitement consiste à répartir l'irradiation de la tumeur sur plusieurs séances : de l'ordre d'une trentaine en radiothérapie conventionnelle (variable selon le type de tumeurs), ou d'une dizaine en hadronthérapie. Dans tous les cas, il faut que le personnel doit donc assurer la reproductibilité du positionnement à chaque séance.

Lors d'une séance de traitement, les incidents potentiels peuvent être classés en 4 types principaux :

1. Identification du patient.
2. Position du patient et de la tumeur.
3. Paramètres d'un ou plusieurs faisceaux (angle, forme, etc.).
4. Valeur de la dose délivrée sur la zone irradiée du patient.

La plupart des incidents des trois premières classes peut être détecté à l'aide d'un système visuel par vidéo. La Fig. 1.3 illustre la possibilité de visualiser ces incidents à l'aide d'une opération de soustraction entre deux images de la même scène, acquises à différents instants. Bien que l'analyse de ces différences au niveau image et la quantification de la différence des positions des éléments entre les deux images sont des tâches complexes, nous sommes très confiants que l'utilisation de techniques de vision par ordinateur et l'exploitation des informations *a priori* dont on dispose sur la scène permettra de répondre aux besoins de l'Assurance Qualité (AQ) en radiothérapie. C'est ce que nous allons essayer de démontrer.

1.2 Besoins

La sécurité lors de séances de traitement en radiothérapie est définie selon deux critères majeurs :

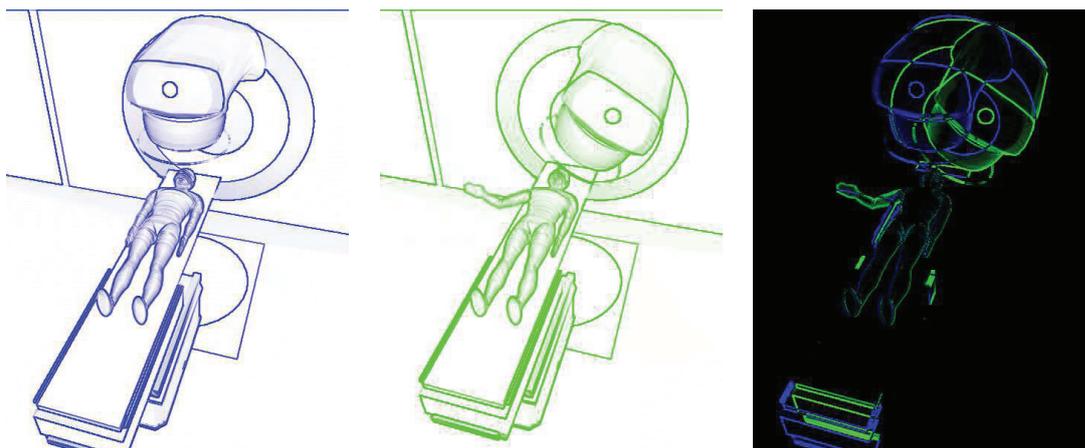


FIGURE 1.3: Exemple d'erreurs pouvant être détectées par l'utilisation d'images vidéo : différences importantes des positions du bras d'irradiation et de la table, fort déplacement du bras du patient, et léger déplacement latéral du corps du patient (5mm).

1. Le déroulement de la séance doit être conforme à ce qui a été planifié et prescrit.
2. Les consignes de sécurité liées à ce type de traitement doivent être respectées.

Après que les radiothérapeutes et les médecins ont défini le plan de traitement, c'est au tour du manipulateur de prendre en charge le patient et de suivre les consignes indiquées. Plusieurs facteurs peuvent être à l'origine d'une erreur à ce niveau, mais les erreurs humaines sont les plus difficiles à détecter. Les outils d'assurance qualité actuellement présents dans les centres de radiothérapie et prenant en charge le facteur humain, ne permettent que de faire une analyse *a posteriori* basée sur le retour d'expérience du personnel médical. Ceci est une tâche supplémentaire et soumise à l'appréciation de ce qui est ou n'est pas une erreur. Ces outils sont focalisés sur un élément (patient ou machine, principalement), et ne prennent pas en compte la globalité des éléments présents dans la salle ni leurs interactions potentielles.

Le manipulateur a besoin d'être assisté dans l'aspect visuel de la surveillance du traitement car il ne peut regarder en continu ce qui se passe à l'intérieur de la salle (Fig. 1.4) Il devrait pouvoir d'être alerté lorsqu'une erreur a lieu pendant le déroulement de la séance afin de limiter les conséquences (notamment au niveau du patient), voire être alerté avant que l'erreur ne se produise. Dans certaines circonstances, quand une erreur a été détectée après une ou plusieurs séances, le radiothérapeute ou le médecin peuvent également avoir besoin de connaître le déroulement des séances précédentes.

1.2.1 Contraintes

Plusieurs contraintes ont été posées, par le personnel médical d'encadrement notamment, dans le cadre de ce projet :

- **Assister le manipulateur sans lui demander d'effort supplémentaire important** : comme on peut le voir sur la Fig. 1.4, le manipulateur a déjà plusieurs systèmes au pupitre (salle de contrôle) avec lesquels il doit interagir pour la délivrance du traitement, en plus des autres tâches qu'il doit réaliser.



FIGURE 1.4: Exemple de poste de contrôle d'une salle de traitement (Centre Léon Bérard à Lyon [7]). On remarque principalement dans cette image que le manipulateur doit travailler avec plusieurs écrans pendant le traitement. Vert : écrans connectés aux caméras à l'intérieur de la salle. Jaune : système d'information, de contrôle (*Record and Verify*), et de pilotage des machines. Blue : autres systèmes pour le positionnement du patient (partie logicielle).

- **Ne pas ajouter de système matériel complexe et encombrant dans la salle de traitement** : notamment parce que les éléments présents prennent une place importante et que leur manipulation est complexe.
- **Ne pas ajouter de matériel supplémentaire qui puisse gêner le patient ou introduire des erreurs potentielles sur son positionnement** : par exemple, des marqueurs collés à la peau, des moyens de contention (déjà utilisés), etc..

1.3 Objectifs

L'objectif général est de développer un système de contrôle externe global par vidéo qui exploite les informations *a priori* contenues dans le dossier du patient. Ce système est complémentaire des différents outils d'assurance qualité déjà présents, et sa principal contribution et aspect novateur est l'utilisation de la vidéosurveillance pour l'analyse du déroulement de la séance et la détection d'erreurs.

1.3.1 Problématiques à considérer

Pour atteindre cet objectif, nous avons listé toutes les problématiques liées aux besoins exprimés par le personnel médical du Centre Léon Bérard [7] et pertinentes à la vidéo-surveillance. Ces problématiques seront rappelées tout au long du manuscrit et permettront d'établir un bilan dans la conclusion au chapitre 8.

Problématiques Générales

- G.1 Validation de la conformité du déroulement d'une séance de traitement par rapport aux informations contenues dans le dossier médical du patient, et surtout celles définies dans le plan de traitement.
- G.2 Validation de la conformité par rapport aux consignes liées au traitement par radiothérapie externe (deux aspects : contexte médical et environnement de radioprotection).
- G.3 Assistance au manipulateur lors du positionnement du patient et lors de l'irradiation de la tumeur, notamment par le contrôle du déroulement de la séance de traitement.
- G.4 Production d'une trace informatisée du déroulement d'une séance de traitement.

Vision par ordinateur

- VO.1 Acquisition synchronisée d'images à partir d'un système composé de plusieurs caméras.
- VO.2 Analyse du contenu d'images provenant d'un et de plusieurs flux vidéo (séquence d'images).
- VO.3 Reconstruction d'informations 3D à partir des informations 2D des images vidéo de plusieurs caméras.
- VO.4 Détection et suivi 3D de mouvements à l'aide des images vidéo avec fond dynamique².
- VO.5 Suivi 3D simultané du mouvement de plusieurs objets.

Surveillance

- S.1 Détection et reconnaissance d'événements dans un milieu contrôlé à partir des informations 2D (images vidéo) et 3D reconstruites.
- S.2 Détection et reconnaissance de scénarios à l'aide d'une base de scénarios prédéfinis.

Modélisation

- M.1 Création d'un modèle numérique de la salle de traitement et de son contenu potentiel (individus et objets) aussi proche que possible de la réalité. Le modèle doit permettre :
 - M.1.a l'intégration des informations contenues dans le dossier du patient et l'accès à ces informations,
 - M.1.b la gestion et la corrélation d'informations de natures différentes,
 - M.1.c l'interaction entre plusieurs entités : homme-homme, homme-objet et/ou objet-objet,

2. déplacements continu des éléments dans l'arrière plan de la scène, autres que les éléments d'intérêt.

M.1.d l'ajout de nouvelles informations et la modification des informations déjà présentes.

Assurance qualité en radiothérapie

- AQ.1 Modélisation du processus d'une séance de traitement.
- AQ.2 Détection et identification des erreurs lors du déroulement d'une séance de traitement.
- AQ.3 Analyse et contrôle des erreurs détectées et identifiées (gestion et maîtrise du risque).
- AQ.4 Identification et évaluation des tolérances aux erreurs en fonction de la précision du système développé et des contraintes établies pour ce type de traitement.

Optimisation

- O.1 Exploitation des informations *a priori* pour réduire la complexité en calcul de différents algorithmes.

1.4 Contextes académique et industriel

Le Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS) [8] travaille dans le domaine médical, entre autres, depuis plusieurs années en proposant des solutions à travers ses compétences en informatique autour de deux thématiques : « Image » et « Données, Connaissances, Services ». Ce laboratoire regroupe environ 280 personnes, dont près de 110 chercheurs et enseignants-chercheurs, avec cinq tutelles : l'Université Claude Bernard Lyon 1 (UCBL), l'INSA de Lyon, l'Ecole Centrale de Lyon, l'Université Lumière Lyon 2, et le Centre National de la Recherche Scientifique (CNRS).

J'ai intégré le laboratoire LIRIS en avril 2008, et plus particulièrement l'équipe SAARA³, pour la réalisation d'un stage de master à la fois recherche et professionnel. Les travaux de l'équipe SAARA s'articulent autour de la réalité augmentée avec la génération d'animations par la création de modèles physiques pour des simulations réalistes, et à partir de l'analyse de mouvements issue des informations acquises par vidéo.

L'objectif de mon stage était de démontrer la faisabilité d'un système de suivi des éléments (individus et objets) présents dans une salle de traitement par radiothérapie [PSM09]. Les résultats ont été présentés au Dr. Hanna KAFROUNI, PDG de l'entreprise DOSIsoft [9], qui a souhaité s'engager dans un travail de thèse par le financement et la participation à l'encadrement dans le cadre d'une convention CIFRE avec le LIRIS.

DOSIsoft est une Société Anonyme créée par l'Institut Gustave Roussy et l'Institut Curie. Cette société assure le développement, la commercialisation et la maintenance de produits informatiques innovants dédiés au domaine de la radiothérapie.

3. Simulation, Analyse, Animation pour la Réalité Augmentée.

Elle est lauréate du concours 2002 du Ministère de la Recherche pour la création des entreprises de technologies innovantes. Afin d'intégrer les innovations thérapeutiques en radiothérapie, les exigences du marché en matière de normalisation (ISO⁴, marquages CE⁵ et FDA⁶) et les évolutions technologiques, DOSIsoft fournit un important investissement en matière de Recherche et de Développement (R&D). Les travaux de R&D sont renforcés grâce à des partenariats avec des institutions de recherche (INRIA, CEA, UCBL et Centres Hospitaliers de référence en France et à l'étranger) permettant ainsi un transfert technologique et une valorisation de leurs travaux de recherche.

Cette thèse s'inscrit dans le cadre d'un partenariat entre le Centre Léon Bérard (CLB) [7] et la société DOSIsoft, créé autour du projet d'Innovation Stratégique Industrielle (ISI) INSPIRA (INformatique pour la Sûreté des Procédés et Installations en RAdiothérapie). Ce projet devrait permettre de développer et de commercialiser des produits logiciels et matériels pour améliorer et renforcer la sécurité et la qualité des traitements en radiothérapie [10]. Deux physiciennes du CLB, Mmes Chantal GINESTET et Pauline DUPUIS, ont participé aux travaux dans le cadre de ma thèse en me faisant profiter de leur expertise, de leur avis et de leurs conseils.

Pendant les six derniers mois, Anthony PEREIRA a participé à ces travaux dans le cadre d'un stage de master, financé par DOSIsoft. Il a eu une forte implication dans la conception et le développement des parties du module de vidéosurveillance présentées dans les chapitres 5 et 6.

4. International Organization for Standardization.

5. Conformité Européenne.

6. Food and Drug Administration, en français : Agence Fédérale américaine des produits alimentaires et médicamenteux.

Chapitre 2

État de l'art

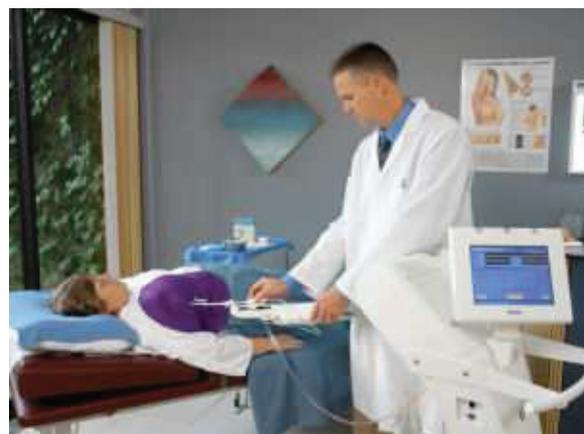
La recherche bibliographique présentée dans ce chapitre est axée sur les problématiques présentées à la section 1.3.1. Une présentation du traitement des cancers par radiothérapie externe est d'abord nécessaire afin de mieux cerner le contexte.

2.1 Radiothérapie et imagerie médicale

En 1895, Wilhelm Conrad Röntgen est la première personne à mettre en évidence un nouveau type de rayon invisible : les rayons X. Au début du siècle dernier, la découverte des propriétés destructrices de ces rayons envers les cellules biologiques, dans plusieurs expériences, donne naissance à la radiothérapie. Cette technique de traitement du cancer consiste à délivrer au volume tumoral une dose de rayonnement létale pour les cellules cancéreuses, mais également toxique pour les cellules saines. On parle de curiethérapie lorsque la dose d'irradiation est délivrée depuis l'intérieur du patient, en implantant directement des sources radioactives au sein de la tumeur à l'aide des « cathéters » (tubes minces, creux et flexibles) ou d'aiguilles. En radiothérapie externe, la source est située à l'extérieur du patient et dirigée vers la zone tumorale, le plus couramment à l'aide d'un accélérateur de particules.



(a) Radiothérapie – Varian [11]



(b) Curiothérapie – Xofig [12]

FIGURE 2.1: Exemples de systèmes de thérapie par irradiation. (a) externe, (b) interne.

Pour l'irradiation de la tumeur, les particules les plus fréquemment utilisés sont :

Les électrons dont les propriétés balistiques permettent de délivrer la dose dès les premiers millimètres de tissus traversés. Leur utilisation est appropriée pour la radiothérapie externe lorsque la zone tumorale est proche à la peau.

Les photons dont les propriétés balistiques permettent de traverser facilement la matière, et utilisés pour traiter des zones tumorales profondes. Un de leurs principaux inconvénients est que le dépôt de dose est réalisé sur tous les organes traversés, d'où le besoin de multiplier les angles d'incidence pour l'irradiation de la zone tumorale pour protéger les tissus sains.

Les protons et les hadrons dont les propriétés balistiques permettent de délivrer une dose plus ponctuelle (par rapport aux électrons et aux photons), et de mieux concentrer la dose sur la zone tumorale et de mieux épargner les tissus sains. Ce sont des techniques spécifiques de la radiothérapie, appelées protonthérapie et hadronthérapie. La précision du dépôt de dose obtenue avec ces particules demande une meilleure prise en compte de mouvements (internes et externes) du patient et ainsi que les tissus traversés.

L'une des grandes évolutions de la radiothérapie a été l'intégration de l'informatique pour la simulation des interactions des particules en fonction des tissus traversés, en utilisant les informations contenues dans des images scanner du patient.

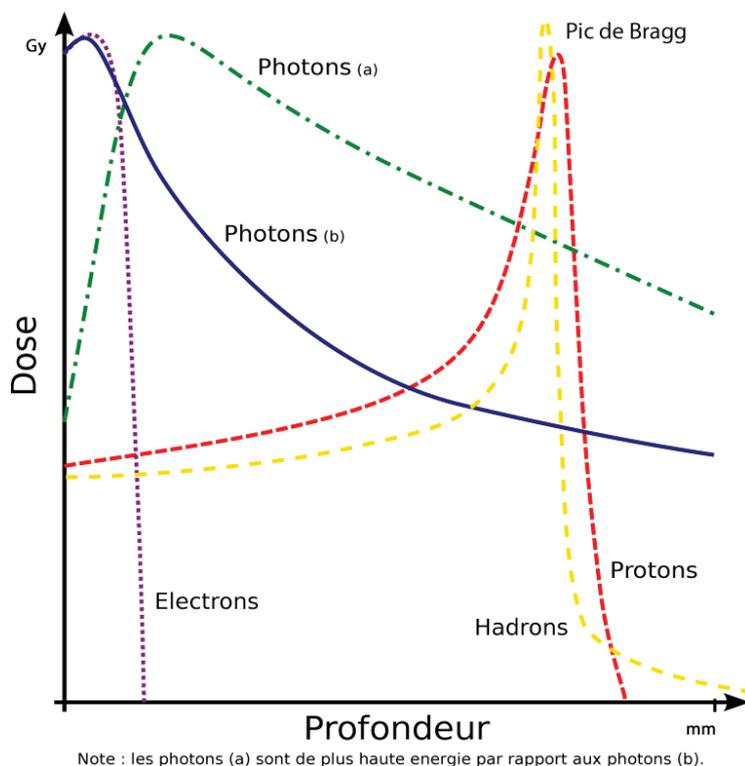


FIGURE 2.2: Dépôt de dose par rapport à la profondeur traversée dans un milieu homogène.

Dans notre cadre de travail, les échanges et les testes réalisés dans le Centre Léon Bérard se faisaient autour de traitements par radiothérapie classique (photons). Les différentes descriptions présentées dans ce document se focalisent sur ce type de traitement. Cependant, nous avons pris en compte les différents types de traitements en radiothérapie (plus ou moins complexes) lors de la conception et du

développement du module de vidéosurveillance, et nous avons essayé de rester le plus générique possible afin de pouvoir répondre aux besoins spécifiques dans une version plus avancée du module.

Trois groupes de professionnels sont en général nécessaires pour la réussite et la sécurisation du traitement [mdls09] :

Le radiothérapeute est le médecin spécialisé en cancérologie et en radiothérapie.

Son rôle inclut :

- l'explication et les conseils au patient sur les options de traitement, et les potentiels effets secondaires à court ou long terme,
- le contourage¹ des organes à risque et des volumes cibles dans le SPT,
- le choix du plan de traitement et de la technique qui sera utilisée,
- le suivi du patient depuis le diagnostic, pendant le traitement, et après la fin du traitement.

Le physicien médical (ou radiophysicien) est le scientifique spécialisé dans les applications médicales des rayonnements (radiothérapie, médecine nucléaire, radiodiagnostic et radioprotection). Son rôle inclut :

- la spécification du matériel utilisé dans l'étude d'imagerie médicale et pendant le traitement (calculs du blindage, choix des paramètres de l'accélérateur, etc.),
- la mise en service et l'assurance qualité des moyens de diagnostic, des équipements et des logiciels pour la planification et pour le traitement,
- la réalisation et la validation des plans de traitement,
- l'exécution, la mesure et l'analyse des faisceaux (il est garant de l'adéquation entre la dose prescrite la dose délivrée).

Le manipulateur d'électroradiologie médicale (ou manipulateur) est un professionnel de santé à la fois soignant et médico-technique. Son rôle inclut :

- l'information, le soutien et le suivi des patients pendant le traitement,
- la réalisation de l'acquisition scanner pour la planification du traitement,
- le repositionnement du patient à chaque début de séance,
- la vérification et la délivrance du traitement à chaque séance.

Un transfert de tâches peut se faire entre les différents professionnels dans certains cas d'organisation spécifique à un service de radiothérapie.

2.1.1 Techniques d'imagerie

Avant d'irradier le patient, il est nécessaire d'obtenir des informations détaillées, et exploitables dosimétriquement, de la région anatomique à traiter. Ces informations sont obtenues par un examen composé en général d'une succession d'images correspondant à des coupes régulièrement espacées, qui vont constituer une représentation 3D de l'intérieur du patient. Les coupes sont réalisées dans un même plan axial car le patient est généralement allongé sur une table qui fait partie intégrante du système d'acquisition. À partir des informations acquises, il est possible de reconstruire des coupes dans n'importe quel plan, afin d'obtenir ultérieurement une

1. marquage avec des traits et des lignes les contours d'une forme.

distribution de dose 3D. La distance entre les coupes et le rapport pixel/millimètres sont les principales contraintes pour l'exploitation de l'examen.

L'examen au scanner est toujours utilisé pour la planification du traitement :

La Tomodensitométrie (TDM) mesure l'absorption des rayons X par les tissus du patient (les tissus durs, notamment les os, sont visibles sur ce type d'examen). Il est composé d'un tube à rayons X et de détecteurs qui tournent autour du patient pour créer des coupes, en mesurant l'absorption des éléments du corps du patient qui se trouvent sur l'axe tube/détecteur. Cette technique est également appelée Scanographie, CT-scan (CT pour Computed Tomography – en français : Tomographie par Ordinateur), ou Scanner Tomodensitométrique pour l'appareil. Les modèles de dernière génération ajoutent un déplacement de la table, synchronisé aux rotations du tube, pour permettre une acquisition hélicoïdale volumique (3D).

D'autres types d'examen peuvent apporter, après fusion avec l'examen au scanner, des informations complémentaires permettant de préciser le contourage des volumes :

L'Imagerie par Résonance Magnétique (IRM) exploite les propriétés de relaxation magnétique de certains noyaux atomiques des cellules du patient, comme les atomes d'hydrogène, présents en grande quantité dans le corps humain. Il en existe plusieurs types, dont l'IRM anatomique (IRMa), qui permet d'obtenir des images en haute résolution spatiale de l'anatomie du patient. C'est la technique la plus couramment utilisée en radiothérapie pour compléter le scanner et pouvoir visualiser les tissus mous. Une autre application du phénomène de résonance magnétique est la Spectrométrie de Résonance Magnétique (SRM). Cette technique permet d'obtenir des images métaboliques du cerveau et de certains autres organes afin de visualiser le fonctionnement (activité cérébrale par exemple) et les réactions bio-chimiques qui se produisent dans les tissus et les organes. Comme pour le scanner, il est possible aujourd'hui d'obtenir une acquisition volumique.

La Tomographie par Émission de Positrons (TEP) mesure la distribution volumique et temporelle d'un radio-isotope² spécifique, injecté en quantité infime dans le corps du patient et qui se fixent sur les cellules malades. Cette technique utilise une caméra sensible aux photons et permet d'obtenir une cartographie des volumes à traiter.

Il est très important de garantir que ces informations seront conformes à la réalité lors de l'irradiation de la tumeur. Pour ceci, il est généralement convenu que le patient doit avoir exactement la même position lors de l'examen et lors de l'irradiation de la tumeur. Pour assurer cette reproductibilité du repositionnement, des moyens de contention et des protocoles d'acquisition sont utilisés. Ceci sera détaillé à la section 2.2.

2. les radio-isotopes émettent des rayons gamma. La détection de ce rayonnement et le traitement des résultats permettent d'obtenir des informations sur l'anatomie de la personne et sur le fonctionnement de certains organes spécifiques.

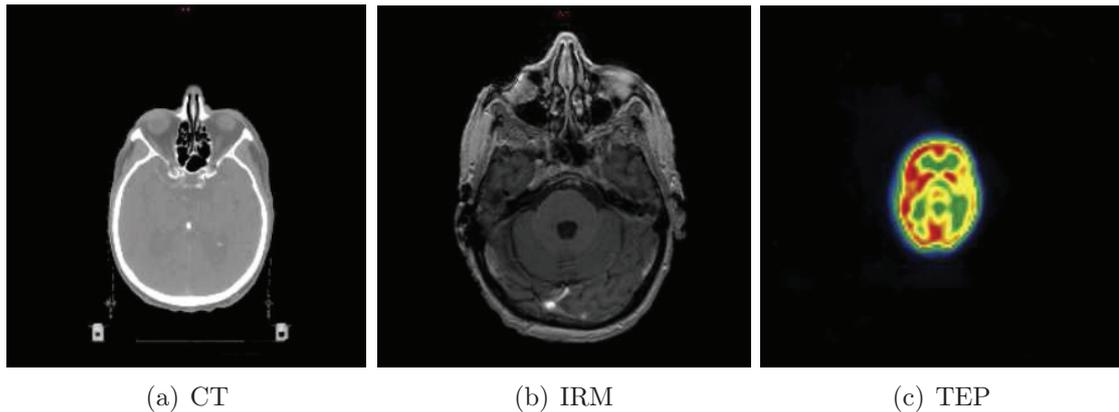


FIGURE 2.3: Exemple de résultats d'examens d'imagerie (coupe au niveau de la tête).

2.1.2 Plan de traitement

À partir de l'examen d'imagerie, le radio-physicien ou le radiothérapeute établit la planification de la dose à délivrer. Le Système de Planification de Traitement (SPT), composé d'outils informatiques de visualisation, simulation et optimisation, permet de déterminer les caractéristiques des faisceaux d'irradiation à appliquer et la distribution de dose prévisionnelle.

La planification du traitement est faite en trois étapes :

Le traitement des images de l'examen d'imagerie consiste à segmenter les volumes d'intérêt (tumeur, organes à risque, etc.). Les contours des volumes sont tracés (de manière automatique, semi-automatique ou manuelle) sur toutes les coupes où ils sont présents. Ensuite, il est possible de réaliser une reconstruction 3D de chacun de ces volumes à partir des contours tracés. À cette étape, il est également possible d'utiliser plusieurs examens d'imagerie de sources différentes, ce qui nécessite le recalage entre les examens avant la segmentation des images.

L'International Commission on Radiation Units and Measurements³ (ICRU) a défini six volumes à prendre en compte, correspondant à des marges de tolérance pour l'irradiation [oRUM93, oRUstIr99] :

Le Volume Tumoral Macroscopique (GTV pour Gross Tumor Volume) est l'étendue macroscopique de la tumeur telle que définie par les examens radiologiques et cliniques.

Le Volume Cible Anatomoclinique (CTV pour Clinical Target Volume) étend le GTV avec une extension microscopique autour de la tumeur (tissus suspects d'un envahissement microscopique prouvé ou potentiel).

Le Volume Tumoral Interne (ITV pour Internal Tumor Volume) englobe le CTV avec une marge supplémentaire pour tenir compte des mouvements physiologiques de la tumeur ou des organes. C'est une marge généralement fondée sur selon l'expérience clinique.

Le Volume Cible Prévisionnel (PTV pour Planning Target Volume) donne une marge de sécurité à l'ITV pour tenir compte des incertitudes dans

3. en français : Commission Internationale des Unités de Mesures.

le positionnement du patient, le système d'immobilisation et les réglages des machines.

Le Volume Traité (TV pour Treated Volume) délimite le volume qui recevra effectivement la dose de traitement, jugé approprié pour atteindre les objectifs de celui-ci.

Le Volume Irradié (IV pour Irradiated Volume) est le volume des tissus qui reçoivent une dose significative par rapport à la tolérance des tissus normaux.

Ces volumes ont été définis pour établir des recommandations par rapport à la prescription et l'évaluation du déroulement des traitements. Il sont également utilisés pour uniformiser les pratiques entre plusieurs centres de radiothérapie faisant partie d'une même étude multicentre.

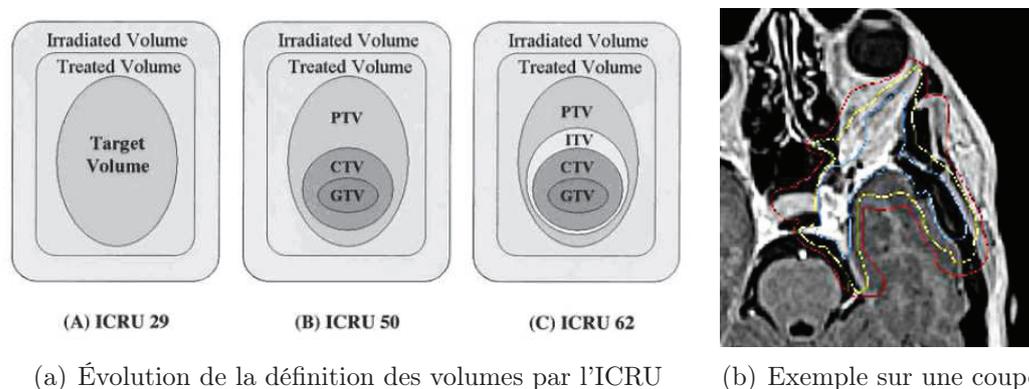


FIGURE 2.4: Volumes de délimitation d'une tumeur.

La détermination de la balistique d'irradiation consiste en la description de l'ensemble des faisceaux. Les paramètres de chaque faisceau, notamment l'incidence (points d'entrée et direction dans le corps du patient), la forme du faisceaux et la configuration d'accessoires, doivent être adaptés afin de maximiser la dose reçue par la tumeur, tout en minimisant celle déposée sur les tissus sains environnants. Une modélisation de la machine est nécessaire afin d'obtenir des simulations les plus proches possible de la réalité. Cette modélisation est basée sur les caractéristiques physiques des particules et de la machine : exploitation des informations sur sa géométrie réelle ainsi que les accessoires (collimateur, applicateur, etc.), les énergies disponibles, etc..

Les paramètres de chaque faisceau sont définis à l'aide du SPT, en vérifiant sur des plans (géométriques) le recouvrement des volumes à irradier par les faisceaux (Fig. 2.5). Ces plans sont parallèles au plan XY dans le repère du faisceau, repère dont l'axe Z définit la direction principale dans laquelle les particules se déplacent.

En 3D, un faisceau a généralement une forme conique à la sortie de la source, et cette forme est modifiée et adaptée au volume à irradier. Dans le cadre de la radiothérapie conventionnelle, des accessoires, appelés collimateurs, permettent de modifier le faisceau selon les axes X et Y . Avec ces accessoires, le faisceau a par défaut une forme pyramidale de base rectangulaire dans le plan XY , mais il est possible d'obtenir d'autres formes pyramidales en modifiant le

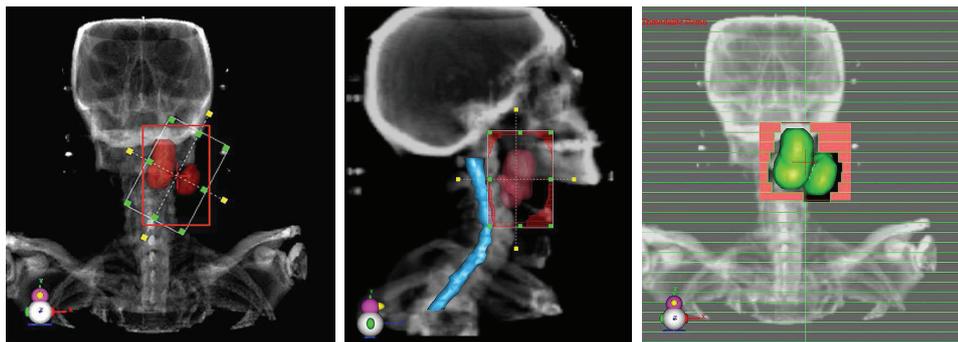


FIGURE 2.5: Captures d'écran du SPT ISOgray®[9] lors de la détermination des faisceaux.

collimateur afin d'obtenir une base polygonale permettant de limiter la taille maximale du champ et de l'adapter à la forme du volume à irradier. Il existe aujourd'hui des collimateurs dynamiques, appelés Collimateurs Multilames (ou MLC pour MultiLeaf Collimator), pouvant changer la forme du faisceau selon les axes X et Y pendant l'irradiation (utilisé notamment dans le cadre de la Radiothérapie d'Intensité Modulé, ou IMRT pour Intensity Modulated Radiation Therapy). D'autres accessoires sont également utilisés pour modifier les paramètres du faisceau en profondeur (filtre en coin, bolus, etc.), c'est-à-dire, selon l'axe Z . Tout l'ensemble permet de délivrer une dose hétérogène adaptée à la tumeur.

L'une des nouvelles techniques de la radiothérapie, l'arc-radiothérapie, permet de délivrer la dose prescrite en faisant tourner la source autour du volume tumoral. L'incidence du faisceau n'est plus un paramètre fixe et des accessoires, notamment le Collimateur Multilames, permettent d'adapter le faisceau à chaque position visitée pendant la rotation. C'est une approche de délivrance de l'IMRT en ajoutant la rotation du bras d'irradiation, qui permet de réduire la dose déposée sur les tissus sains, bien que le volume de tissus sains irradiés soit plus important.

Dans le cas spécifique de l'hadronthérapie, le phénomène du Pic du Bragg des hadrons est exploité, à la place des accessoires utilisés en radiothérapie classique, pour modifier le faisceau selon l'axe Z pour mieux cibler le volume à irradier. Ce phénomène est caractérisé par un pic important de dépôt de dose qui se produit juste avant que les particules ne s'arrêtent (voir Fig. 2.2).

Le calcul de la distribution de dose en 3D. La dose est exprimée en Gray (Gy), 1 Gy correspondant à la quantité d'énergie absorbée par unité de masse irradiée ($Gy = J/Kg$). Une simulation numérique permet de vérifier que les paramètres définis pour l'ensemble des faisceaux respectent les contraintes imposées (dose minimale nécessaire dans la tumeur ciblée, dose maximale tolérée dans les organes à risque, etc.).

Le fractionnement, qui consiste à répartir la dose totale sur plusieurs séances, favorise la réparation des lésions induites par les rayonnements ionisants des cellules saines, plus rapide que celle des cellules cancéreuses à cause de leur moindre radiosensibilité (ou de leur meilleure capacité de réparation). Le fractionnement classique est de 5 séances par semaine, avec une dose par fraction égale à 2Gy. L'intervalle entre deux fractions et l'étalement de la dose (nombre

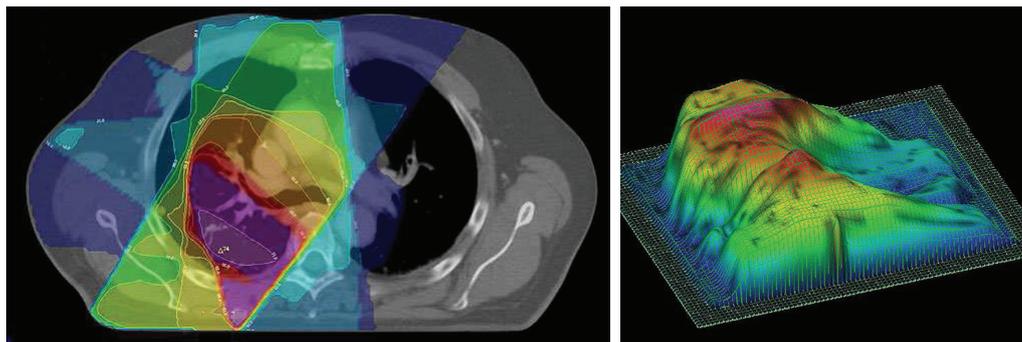


FIGURE 2.6: Captures d'écran du SPT ISOgray®[9] lors de l'étude de dosimétrie.

de jours entre la première et la dernière séance de traitement) ont un rôle important dans le traitement, notamment sur la probabilité de contrôle tumoral et de la toxicité sur les tissus sains [MM05]. L'hadronthérapie permet de réduire le nombre de séances par un facteur important (de l'ordre de 3 à 5, suivant les indications). On parle alors d'hypofractionnement.

2.1.3 Déroulement des séances

Pendant une séance de traitement, le manipulateur est la personne qui prend en charge le patient. Une séance de traitement commence par le positionnement du patient. Comme indiqué à la section 2.1.1, le but est de reproduire la position de référence que le patient avait lors du scanner pour la dosimétrie. Cette tâche se fait généralement en deux étapes :

1. le patient est positionné en fonction des informations de référence, obtenues à partir du scanner et éventuellement marquées sur la peau du patient et/ou à la surface des systèmes de contention,
2. des techniques d'imagerie par rayons X sont ensuite utilisées pour obtenir des informations internes du patient afin de contrôler le positionnement, notamment du volume à irradier (on parle alors de la Radiothérapie Guidée par l'Image, ou IGRT pour Image-Guided Radiation Therapy). Si nécessaire, des corrections sont appliquées à la position du patient.

Les moyens et systèmes utilisés pour cette tâche de positionnement seront détaillés aux sections 2.2.1 et 2.2.2.

Quand le patient est bien positionné, le personnel médical doit sortir de la salle de traitement pour que l'irradiation puisse débuter, pour éviter d'être irradiés (consigne de radioprotection). Le patient reste donc seul dans la salle, porte fermée, jusqu'à la fin de la séance. Le manipulateur peut commander l'appareil de traitement depuis l'extérieur de la salle (poste de contrôle – ou pupitre) en disposant d'une vue de l'intérieur à l'aide des caméras installées dans la salle de traitement. Une fois le manipulateur à l'extérieur, la dose prescrite est délivrée. Pendant l'irradiation, le patient pourrait effectuer des mouvements susceptibles de compromettre la conformité du positionnement. Des systèmes, qui seront également détaillés à la section 2.2.2, permettent de contrôler la position du patient pendant l'irradiation.

Une fois que toutes les séances planifiées ont été réalisées, un suivi médical est nécessaire afin de contrôler les effets secondaires éventuels du traitement et/ou une

possible récurrence de la maladie. Le patient doit se rendre à des consultations avec son médecin, à une fréquence bien déterminée, pendant lesquelles il répond à des questionnaires et subit des examens physiques.

2.2 Assistance au positionnement du patient

Lors de la délivrance du traitement, le manipulateur doit pouvoir assurer qu'il reproduit bien ce qui a été planifié. La tâche de positionnement doit être réalisée de telle sorte que le volume traité coïncide au mieux avec le volume cible théorique. De nombreuses études ont cherché à évaluer l'importance des erreurs de positionnement du patient, et des travaux déjà anciens ont démontré qu'un (re)positionnement imparfait pouvait conduire à une dégradation des résultats thérapeutiques en termes de contrôle local et de survie [XLD⁺00, CSM⁺03, Web06, SSW⁺10].

Au cours du temps, les traitements par radiothérapie et leur planification sont devenus complexes et très précis avec l'arrivée des nouvelles technologies. Ainsi, la qualité du positionnement est de plus en plus critique (en considérant toujours les mouvements internes et externes : mouvements physiologiques, comme la respiration, ou mouvements « volontaires » du patient de l'un de ses membres ou de tout son corps). Devant cette problématique, deux solutions ont été proposées pour améliorer le positionnement du patient :

1. L'immobilisation du patient pendant la délivrance du traitement (irradiation).
2. Le contrôle régulier du positionnement du patient et la correction de la position si nécessaire.



(a) Ingenuity CT System [13]

(b) HexaPODTMevo RT System [14]

FIGURE 2.7: Exemples de tables pour l'acquisition scanner (gauche) et le traitement (droite).

2.2.1 Systèmes d'immobilisation

Un système de contention permet d'immobiliser le patient, principalement au niveau de la région anatomique à traiter (tête, thorax, bassin, etc.), et de le (re)positionner de la même manière à chaque séance. Cependant, l'efficacité n'est pas absolue car la morphologie du patient peut varier en cours des séances [PYSaVH⁺96].

Les contentions, qui peuvent être attachées à la table, sont confectionnées par le manipulateur avant l'étude d'imagerie et utilisées tout au long du traitement. Le



FIGURE 2.8: Exemples de systèmes de contention.

confort du patient est un facteur important pour assurer l'immobilisation afin de réduire le risque que le patient soit dans une position inconfortable et bouge, soit parce que le traitement est long, soit parce qu'il ne peut pas supporter la contention. Les systèmes les plus couramment utilisés sont ceux qui s'adaptent à la forme du patient, comme les moules ou les masques de contention, souvent composés d'un matériau thermoformable à basse température et qui gardent la forme du patient une fois que le matériau s'est refroidi. Ces contentions personnalisées sont confectionnées par les manipulateurs. La contention est un outil impératif pour le traitement, bien qu'il ne représente qu'un des garde-fous dans la chaîne d'assurance qualité.

2.2.2 Systèmes de contrôle lors du positionnement

Une fois que le patient est placé sur la table à l'aide des contentions, le déplacement de la table permet de placer le volume à irradier dans la position planifiée. Il est important de faire remarquer que l'écart de la position de la table entre deux séances est négligeable, et que seul l'écart de la position du patient à une position de référence est pris en compte. Autrement dit, la référence de la position du patient est définie par rapport au référentiel global et non par rapport au référentiel de la table. De plus, le positionnement du patient assure la position du volume cible de manière indirecte, et les mouvements potentiels de la tumeur à l'intérieur du patient sont compensés par les marges de tolérance. L'isocentre, qui correspond au point d'intersection des axes de l'ensemble des faisceaux de traitement, est le centre du repère utilisé pour positionner le patient (voir Fig. 2.9).

Systèmes basés sur des repères externes : une première phase dans le positionnement consiste à utiliser des repères externes afin de réduire le recours à des techniques d'imagerie. Le grand intérêt des systèmes externes est leur aspect temps réel et le fait de ne pas délivrer de dose. Les méthodes de vision par ordinateur utilisées dans ces systèmes seront détaillées à la section 2.4.

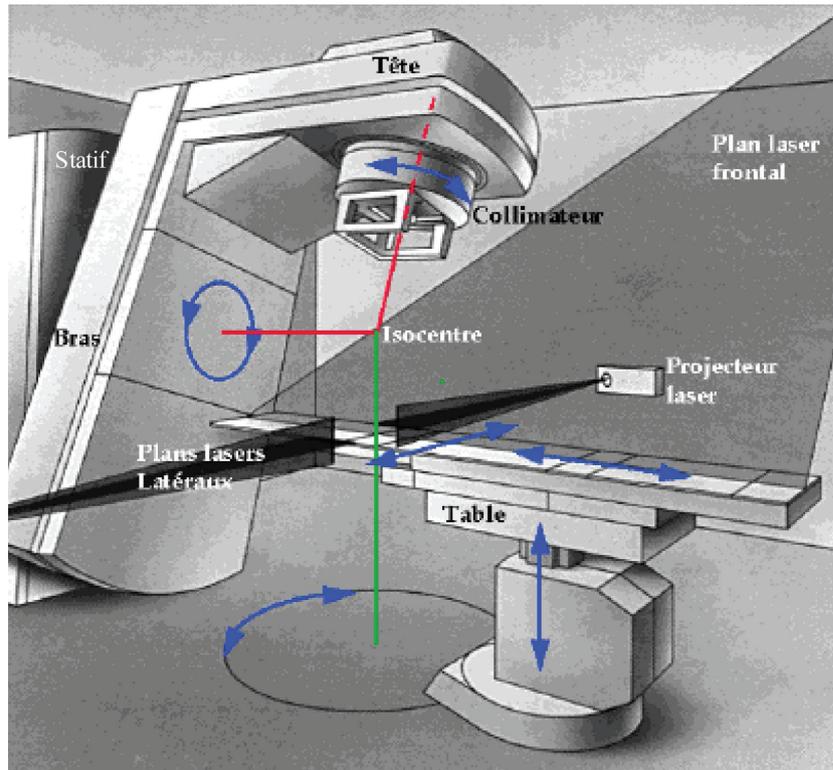


FIGURE 2.9: Bras isocentrique et ses degrés de liberté (doc. SIEMENS).

Les lasers de positionnement sont utilisés très régulièrement. Trois lasers fixes permettent de définir 3 plans orthogonaux (un pour chaque dimension de l'espace) dont l'intersection est à l'isocentre. Les lasers sont utilisés par le manipulateur pour tracer des marqueurs sur la peau du patient ou sur les contentions lors de la première séance à l'aide d'un feutre (durable mais effaçable). Il est également possible, au lieu de tracer de tels marqueurs, d'utiliser des tatouages⁴ ou des gommettes avec des marqueurs imprimés. Le positionnement du patient dans les séances suivantes consiste à aligner ces marqueurs par rapport aux lasers.

Le faisceau lumineux constitue un système analogue aux lasers. Il s'agit d'une source lumineuse, contenue dans le bras d'irradiation, qui permet de simuler géométriquement l'un des faisceaux qui seront délivrés au patient. Des marqueurs sont tracés le long des contours de la projection de la source lumineuse sur la peau du patient.

Les systèmes optiques avec marqueurs permettent de calculer la position 3D des repères externes en temps réel. Les configurations peuvent varier en fonction de la technologie utilisée (p. ex. infrarouges, points de couleurs, motifs, etc.) mais le principe reste le même : les marqueurs sont enregistrés grâce à un système de caméras, dont les paramètres de calibrage (intrinsèques et extrinsèques) permettent de calculer la position 3D de chaque marqueur en utilisant une méthode de reconstruction visuelle par ordinateur (p. ex. stéréovision, recalage 3D, etc.). Reposant sur ce

4. de par le caractère « indélébile » du tatouage, on essaie souvent de réaliser des points ou de petites croix, plus faciles à dissimuler par le patient.

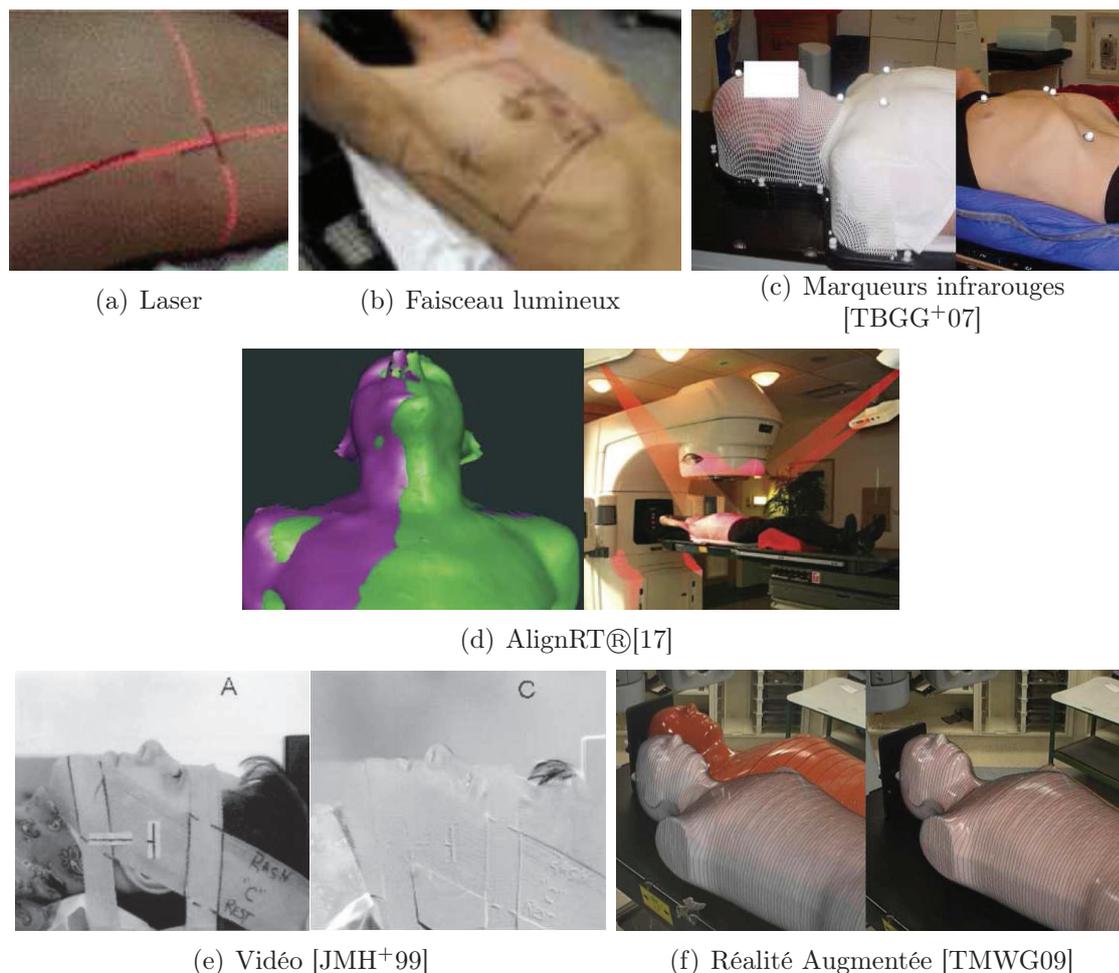


FIGURE 2.10: Exemples de systèmes externes de positionnement.

principe (infrarouge + stéréovision), le système Polaris [18] est beaucoup utilisé dans le domaine médical [PMF⁺07]. D'autres exemples, spécifiques à la radiothérapie, sont présentés dans [BFOP00, SBR⁺06, TWW⁺06].

Les systèmes optiques avec projection de motif permettent la reconstruction d'un profil ou d'une surface correspondant à une partie de la peau du patient (de préférence sur la zone à irradier) par une approche optique (stéréovision et/ou lumière structurée). Les informations 3D acquises sont ensuite comparées aux informations simulées, qui ont été calculées sur les images de l'étude d'imagerie (profil ou surface en position idéale). Le système le plus connu mettant en place ce principe (avec reconstruction de surface) est AlignRT®[17] [SMH⁺03].

La vidéo est utilisée comme système optique sans marqueur qui repose sur le calcul de différence entre deux images : une image de référence (prise après le positionnement de la première séance) et une image provenant du flux vidéo. L'utilisation de la vidéo donne des résultats avec une précision proche des autres systèmes, avec l'ajout de la visualisation des différences non liées directement au positionnement du patient (par exemple, manque d'un accessoire) [CBV⁺75, MRH⁺97, JMH⁺99, YSB02, PFB⁺03]. Talbot et al. [TMWG09] ont proposé un système exploitant la réalité augmentée

pour assister le positionnement du patient. Ce système utilise les contours externes⁵ du patient pour créer une image virtuelle équivalente, considérée comme l'image de référence. Pendant que le manipulateur réalise le positionnement, il peut contrôler sur un écran la superposition de l'image virtuelle sur les images réelles acquises et ainsi obtenir une confirmation visuelle du positionnement (pas de calcul de différence ou d'analyse automatisée).

Systèmes basés sur des repères internes : les repères externes sont imparfaits puisqu'ils sont mobiles par rapport aux structures internes à irradier, et peu précis par rapport aux besoins dans le cas de repères tracés (tâche manuelle). Une deuxième phase permet de contrôler la position du volume à irradier à l'aide d'une technique d'imagerie (une de celles utilisées lors de l'étude d'imagerie pour la planification – cf. section 2.1.1) et d'apporter une correction si nécessaire. La fusion des résultats de plusieurs méthodes est également possible. Le principe est d'obtenir une représentation radiologique numérique des organes internes du patient et de les comparer ensuite à une représentation acquise et validée précédemment, ou bien au volume de référence (étude d'imagerie).

L'imagerie portale (ou EPID pour Electronic Portal Imaging Device) est l'une des techniques les plus répandues. L'imageur portal est solidaire au bras d'irradiation, et permet de générer une image 2D correspondant au champ d'irradiation, c'est-à-dire à la projection de rayons X issus de l'accélérateur avec une dose faible. Cette image peut ensuite être comparée à une image DRR (Digitally Reconstructed Radiograph⁶), qui est une image générée par ordinateur et reconstruite à partir des informations acquises pendant le scanner pour la dosimétrie (potentiellement sous n'importe quel angle, bien qu'on s'intéresse uniquement aux incidences des faisceaux du plan de traitement)[PST⁺00]. Une variante de l'utilisation de la tomodensitométrie est d'inclure un scanner CT dans la salle de traitement [KOS⁺03].

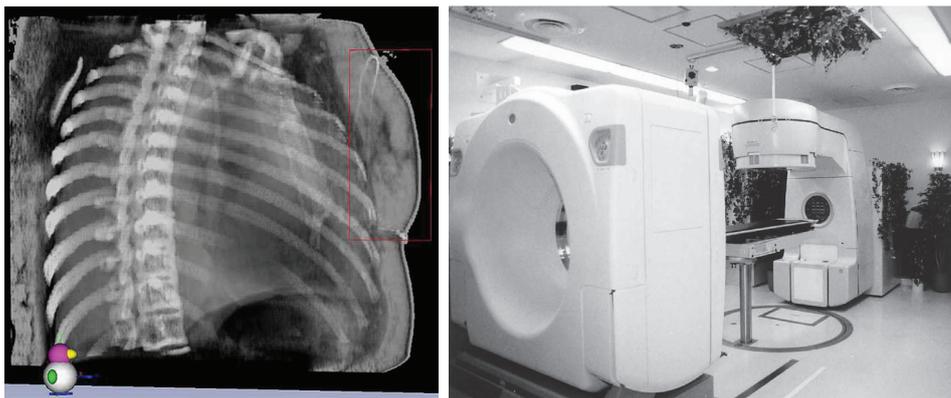
L'imagerie Cone Beam (CBCT pour Cone Beam Computer Tomography) consiste en la réalisation d'une reconstruction volumique à partir des images portales prises lors d'une rotation autour du patient. Deux types d'acquisition existent, en fonction de la source :

1. le MV-CBCT (MV pour Megavoltage) utilise des faisceaux haute énergie, comme les faisceaux utilisés lors du traitement,
2. le kV-CBCT (kV pour Kilovoltage) utilise des faisceaux de plus faible énergie générés par une source de rayons X montée sur le système d'irradiation.

L'imagerie par ultrasons (échographie) est une option non invasive et non irradiante pour le patient, temps réel et moins coûteuse que les autres techniques. Par essence, cette technique ne peut être utilisée que dans les cas où la zone traitée permet une propagation adéquate des ultra-sons

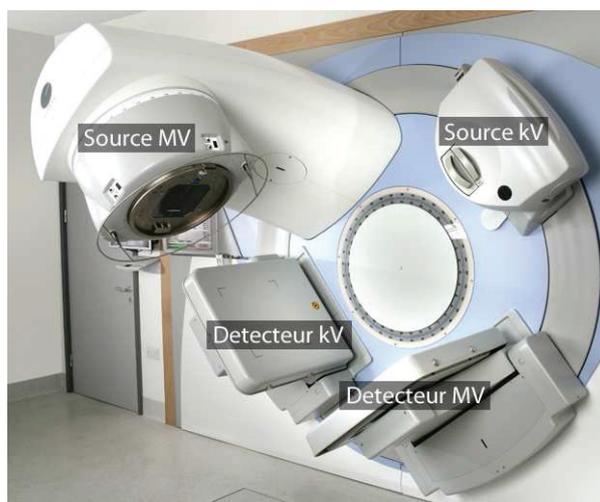
5. surface 3D représentant la peau du patient et obtenue à partir de l'étude d'imagerie.

6. en français : Radiographie Reconstituée Numériquement.



(a) DRR – TPS ISOgray® [9]

(b) Scanner couplé au LINAC [KOS+03]



(c) Imageurs portaux MV et kV couplés au LINAC
– Elekta Synergy® System[14]

FIGURE 2.11: Exemples de systèmes internes de positionnement.

(notamment dans un milieu de tissus mous) [LMP+99, FSS+01, SCB+02, LDL+03].

L'imagerie TEP est utilisée depuis peu dans le positionnement, du fait de sa capacité de calcul de la distribution spatiale et dosimétrique du faisceau en hadronthérapie [Web05].

Si une correction de la position est nécessaire, l'idéal est de la réaliser à l'aide des informations externes et ainsi d'éviter la réutilisation d'une technique d'imagerie. La transformation à réaliser est alors calculée par rapport aux informations internes et ensuite appliquée par rapport aux informations externes. Ceci peut être réalisé à condition que la relation entre les informations internes et externes soit connue. Les marqueurs radio-opaques⁷, implantés à l'intérieur du patient, peuvent être utilisés pour simplifier cette tâche [BLS+95, VPL+97, LDS+02].

7. non transparent aux rayons X ou à d'autres formes de radiation.

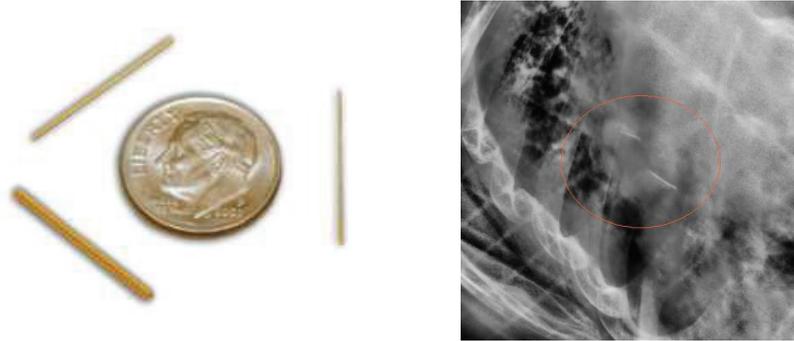


FIGURE 2.12: Exemple de marqueurs radio-opaques.

2.2.3 Systèmes de contrôle lors de l'irradiation

L'immobilisation du patient n'étant pas parfaite, le positionnement du patient n'est valable dans le temps que pendant une durée limitée. Pour un bon déroulement de la séance, il faudrait que cette durée soit au moins égale à la durée de l'irradiation. La mise en place de moyens de contrôle du positionnement pendant la séance permettrait alors de quantifier cette durée de validité, notamment au moyen de la quantification rigoureuse des déplacements du volume cible liés aux processus physiologiques (respiration, digestion, etc.).

De tous les systèmes présentés précédemment, ceux permettant une acquisition temps réel sont généralement utilisés pour la tâche de contrôle du positionnement. Pour les systèmes externes, une corrélation entre les informations internes et externes est nécessaire. À titre d'exemple, des marqueurs radio-opaques et infrarouges peuvent être utilisés comme repères externes [SBR⁺06]. La caractéristique infrarouge des marqueurs permet de suivre leurs positions afin d'en déduire celle des repères internes à chaque instant, à l'aide de la relation entre les repères internes et externes. Cette relation est obtenue en exploitant la caractéristique radio-opaque des repères dans un étude d'imagerie (pouvant comporter plusieurs instants du mouvement), ce qui permet de définir un modèle de comportement (par exemple, mouvement cycliques, linéaire, etc.) [SDV⁺10, SFR⁺11]. Autrement dit, combiner des systèmes externes et internes (représentant la tumeur) permet de modéliser le déplacement de la tumeur en fonction de la position des repères externes (principalement utilisé pour la respiration). Un autre exemple est le système GateRT [17], qui exploite la différence entre les surfaces acquises à différents instants et des seuils de tolérance de différence entre surfaces.

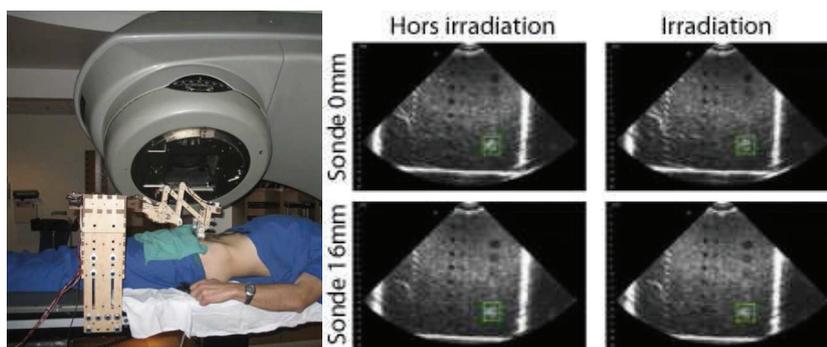
Évidemment, les systèmes internes permettent de calculer avec plus de précision la position de la tumeur [BCP⁺93, BWN⁺05]. Aujourd'hui, on peut retrouver dans les salles de radiothérapie des systèmes de monitoring interne, utilisant par exemple des technologies comme l'imagerie portale [BJS⁺04, VTB⁺06], les ultrasons [FSC⁺04, HME⁺05, BMDT09], ou la radio-identification (RFID pour Radio-frequency identification) [WKP⁺06, MKO⁺08].

Le principal défi de ces systèmes de contrôle est de pouvoir suivre les mouvements internes dans des zones où les mouvements sont plus importants (thorax, prostate, etc.). Cependant, ces systèmes restent focalisés sur la zone à irradier du patient, et ne prennent pas en compte les éléments qui l'entourent (système d'immobilisation, table de traitement, etc.).



(a) GateRT (external) [17]

(b) Calypso®4D (internal) [19]



(c) Ultrasons. À gauche : robot téléguidé contenant la sonde. À droite : images ultrason. [SSH10].

FIGURE 2.13: Exemples de systèmes de monitoring.

2.3 Contrôle en radiothérapie

La radiothérapie est un domaine complexe qui requiert l'utilisation de technologies de pointe. Dans le but de respecter à chaque étape du processus (Fig. 1.1) une tolérance acceptable pour réduire la probabilité d'apparition d'écarts et d'incidents, voire d'accidents, il est impératif que des moyens de contrôle soient mis en place. Ces moyens doivent être adaptés à chaque étape pour assurer la sécurité du traitement et portent sur [Gui00] :

- les normes de sécurité internationales pour la protection du patient et du personnel contre l'irradiation,
- l'utilisation optimale, la précision et la sécurité des équipements,
- le bon usage des installations,
- la précision par rapport les grandeurs dosimétriques et géométriques,
- la planification du traitement pour optimiser la distribution de la dose,
- les situations d'urgence,
- la sécurité et la protection physique des sources d'irradiation.

Mais ce contrôle ne doit pas se limiter aux aspects physiques et techniques du processus de traitement. Il doit également englober toutes les activités du centre de radiothérapie, à partir du moment où un patient entre jusqu'au moment où il le quitte, et il doit se poursuivre pendant toute la période de suivi [mdls09].

2.3.1 Assurance Qualité en radiothérapie

L'Organisation Mondiale de la Santé (OMS) [5] donne la définition suivante :

L'assurance qualité (AQ) en radiothérapie est l'ensemble des procédures qui assurent la cohérence de la prescription médicale, et l'accomplissement en toute sécurité de cette ordonnance, concernant la dose au volume ciblé, avec une dose aussi faible que possible sur les tissus sains, une exposition minimale du personnel et la surveillance adéquate des patients visant à déterminer le résultat final du traitement. [mdls88].

Ces procédures [PA05] :

- permettent de réduire les incertitudes et les erreurs, ce qui améliore la précision et donc les résultats du traitement,
- augmentent la probabilité de reconnaître les accidents et les erreurs qui ont eu lieu, afin de pouvoir les corriger plus tôt et de réduire ainsi leurs conséquences pour le traitement du patient,
- permettent une comparaison fiable des résultats entre les différents centres de radiothérapie, en assurant un traitement plus uniforme et précis. Cela est nécessaire pour les essais cliniques et aussi pour le partage d'expériences cliniques en radiothérapie et leur transfert entre les centres,
- doivent garantir le haut niveau de précision et de cohérence nécessaires pour l'exploitation des avancées technologiques et des nouvelles techniques de traitement,
- doivent garantir la sécurité du patient, ce qui est déjà l'un des objectifs de la radiothérapie, notamment en termes de l'exposition des tissus sains et des organes à risque qui doit être maintenue « aussi basse que raisonnablement possible » (ALARA pour As Low As Reasonably Achievable) et compatible avec la délivrance de la dose requise pour le volume cible.

Les moyens utilisés pour suivre ces procédures sont donc diverses, par exemple : détecteurs pour la dosimétrie *in vivo*, rapports de référence pour l'étalonnage des faisceaux (métrologie des rayonnements ionisants), textes de loi pour le contrôle qualité des appareils de traitement, logiciels pour l'organisation du traitement et du personnel, repères visuels pour le positionnement, etc.

2.3.2 Gestion des risques

Le risque est par définition (ISO 14971⁸) la combinaison de la probabilité d'un événement⁹ et de ses conséquences. Dans la sécurisation des traitements en radiothérapie, le nouvel enjeu consiste à mettre en place des politiques de « défense en profondeur », par le biais dans la mise en place de lignes de défense indépendantes et successives (Fig. 2.14). Ces lignes de défense assurent le respect des limites établies pour les conséquences de tout événement qui puisse se produire. L'ensemble des lignes de défense doit assurer la sécurisation globale du traitement, même lorsqu'une des lignes de défense est compromise ou défaillante. Cet enjeu est principalement focalisé sur les événements dont l'occurrence est incertaine, et dont les conséquences ont un impact négatif ou réduisent le gain et/ou l'efficacité du traitement [dSN11].

La prise de risque fait partie de toute activité humaine et elle est liée à la recherche d'un bénéfice, à condition que l'erreur puisse d'être anticipée et identifiée

8. Dispositifs médicaux – Application de la gestion des risques aux dispositifs médicaux.

9. Un événement est l'occurrence d'un ensemble particulier de circonstances (ISO/CEI 73).

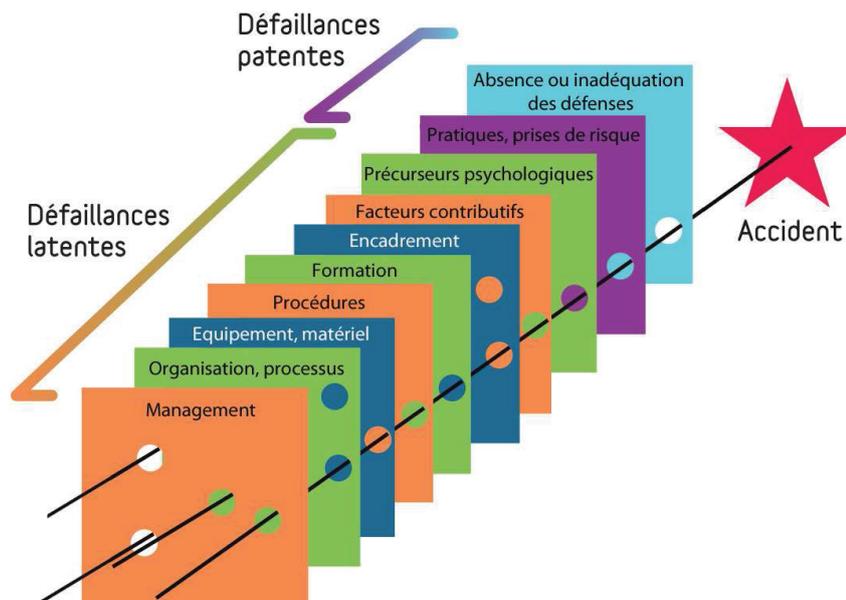


FIGURE 2.14: Sources de défaillances d'une organisation, d'après Reason [Rea97].

afin d'être contrôlée. Une échelle d'acceptabilité, souvent nécessaire, permettra de mesurer une activité en fonction du risque (tolérable, justifiable, etc.) et des résultats (satisfaisant, donnant un gain supplémentaire, etc.) avec la prise en compte de l'acceptabilité des différents acteurs. Ensuite, l'intervention des acteurs peut avoir deux effets opposés : générer le danger et le réduire. À titre d'exemple, environ 80% des erreurs sont détectées par la personne qui les a commises [FL09]. Le facteur humain occupe donc une place centrale dans toute analyse de risque [RK08].

2.3.3 Contrôle du risque humain

Les facteurs organisationnels et humains, en particulier l'insuffisance de ressources humaines et la charge de travail excessive ou inadéquate, et des défauts dans l'éducation et la formation, ont été identifiés comme les principales causes d'événements indésirables en radiothérapie [YBC⁺05, CIR⁺10, dCL11].

L'approche normative a été l'une des premières à être mises en place. Ainsi, des organismes de santé et de radioprotection dans le monde ont établi des consignes et des obligations pour les centres de radiothérapie [RA01, dedeSA03, Uni06, dSN08, dlSedSF09, mdls09]. Cependant, De Jong et al. [dJNR11] ont pu mettre en évidence des écarts importants de ressources humaines (dans le nombre de personnes par profession), notamment au niveau de la responsabilité par discipline (médecin, physicien et manipulateur), dans différents services de radiothérapie dans le monde entier entre 2006 et 2009.

La piste suivie actuellement est de tirer les enseignements des accidents survenus dans le passé pour prévenir les accidents futurs [CBP⁺10]. À titre d'exemple, le système Radiation Oncology Safety Information System¹⁰ (ROSIS) [CCKH10], utilisé au niveau international, permet de rassembler et de partager des informations

10. en français : Système d'Information de Sécurité pour la Radio-Oncologie

sur les incidents, ainsi que le contexte dans lequel ils sont survenus, et les procédures mises en place par les services de radiothérapie pour corriger les failles. En France, des cellules de retour d'expérience (CREx) [LCDD⁺08] ont été implantées dans de nombreux hôpitaux comme outil de gestion des risques par l'enregistrement d'évènements indésirables déclarés, l'analyse des causes de ces événements et la correction de pratique afin qu'ils ne se reproduisent plus. Le système ORION® [DBN⁺11] a été également mis en place afin de faciliter la pratique courante de l'analyse systémique¹¹ des incidents répertoriés, considérés évidemment comme une cause potentielle de danger. Cette analyse permettrait la réalisation d'actions correctives efficaces pour l'hôpital et l'amélioration de la qualité des soins pour les patients.

2.4 Méthodes de suivi d'individus et d'objets

Le suivi de mouvement est un domaine de recherche ancien et encore très actif, avec des nombreuses applications. Les premiers travaux débutent plusieurs années en arrière avec l'envie de connaître la nature du mouvement en lui même. Après l'avènement de la photographie (1876) puis de l'image numérique, l'arrivée des caméras vidéos connectées aux ordinateurs permet, entre autres, d'automatiser la capture et l'analyse des différentes positions d'un mouvement effectué par un humain ou un objet dans une séquence d'images.

L'utilisation de caméras vidéo dans le domaine médical, couplée aux techniques de vision par ordinateur, a été introduite pour l'étude morphologique et physiologique des animaux [Mar87, Muy87] et de l'humain [Joh73]. Leur utilisation s'étend aujourd'hui à la télévigilance médicale [CPV07, MIB⁺09] et à l'assistance du médecin aux gestes chirurgicaux [SAN⁺04, NPS⁺09], entre autres. Cette section se concentre sur les méthodes de vision par ordinateur utilisant des caméras vidéo pouvant répondre à nos besoins, notamment concernant le suivi simultané de plusieurs éléments¹².

Le processus du suivi de mouvement a deux principales composantes :

L'estimation du modèle de l'élément est le processus qui détermine une représentation numérique de la structure géométrique, cinématique et dynamique de l'élément filmé.

L'acquisition de mouvements est le processus qui estime les paramètres de cette représentation à tout instant. Le mouvement de l'élément est décrit par une suite de valeurs des paramètres, acquis dans une durée de temps donnée.

Le suivi de mouvement inclut le procédé de capture des mouvements (motion capture ou mocap) d'un élément réel pour les enregistrer ou les transmettre en temps réel vers d'autres systèmes. On distingue quatre types de ce procédé : optiques, mécaniques, magnétiques et inertiels [Men99, MG01, Pop07]. Dans la suite de ce document nous nous intéresserons aux systèmes optiques, et nous en présenterons quelques-uns.

11. L'analyse systémique consiste dans la modélisation d'un système complexe afin d'identifier les caractéristiques qui permettent de mieux comprendre son fonctionnement, ses interactions avec d'autres systèmes et la manière d'influer dessus.

12. élément : terme qui sera utilisé pour désigner indifféremment les individus et les objets

2.4.1 Représentation numérique de l'élément

L'utilisation d'un modèle numérique permet d'exploiter des informations sur l'élément suivi, et dans certains cas sur la scène, connues *a priori* ou acquises pendant une phase d'apprentissage ou en continu.

Le modèle géométrique correspond à la représentation de la forme de l'élément dans l'espace. Parmi les représentations les plus utilisées on trouve les nuages de points 3D, les maillages (représentation surfacique) et les ensembles de voxels (représentation volumique). Il peut être généré à partir des mesures géométriques de l'élément ou bien obtenu automatiquement à l'aide d'une technique d'acquisition de données et d'une technique de reconstruction. Lorsque des techniques de Conception Assistée par Ordinateur (CAO) sont utilisées, le modèle numérique de l'élément est connu avant même l'existence de celui-ci. Cependant, des différences peuvent surgir entre le modèle original et l'élément créée, et des corrections sont alors nécessaires. Parmi les techniques d'acquisition automatique, on trouve les techniques d'imagerie présentées à la section 2.1.1, les techniques d'acquisition surfacique utilisant le laser [FFG⁺01] ou les techniques de reconstruction à partir de prises de vue multiples de l'élément [MGB05].

Le modèle cinématique correspond à la représentation du mouvement de l'élément indépendamment des causes. Il est défini par une structure contenant un ensemble de paramètres, et des valeurs théoriques possibles pour ces paramètres. L'une des représentations les plus classiques, notamment pour les humanoïdes, est la représentation hiérarchique des différentes parties rigides de l'élément, connectées par des articulations. Dans la formalisation du modèle, les différences sont principalement présentes dans la modélisation des articulations, comme par exemple les angles d'Euler [Gav99, DF01], les quaternions [KM95] ou le principe « axe et angle » très utilisé en robotique [MLS94, BMP04].

Le modèle dynamique correspond à la représentation du mouvement de l'élément dans le temps. Il est souvent basé sur des outils probabilistes et exploité comme outil de prédiction et/ou correction de la localisation de l'élément, afin d'augmenter la robustesse du suivi. Parmi les techniques les plus courantes on trouve les filtres particuliers [DBR00, GSS93], les méthodes de Monte-Carlo par chaînes de Markov [Pie03], le filtre de Kalman [Kal60] et l'algorithme CONDENSATION (CONDitional DENSity PropagATION) [IB98].

2.4.2 Techniques d'acquisition du mouvement

Dans le suivi de mouvement par vidéo, l'idée de base est de détecter et d'identifier les caractéristiques des éléments dans les images. L'exploitation des paramètres du système d'acquisition, notamment les paramètres intrinsèques et extrinsèques des caméras, permet d'obtenir des informations 3D :

- par l'estimation du mouvement des caractéristiques de l'élément dans l'image, pour chaque caméra indépendamment, et le suivi d'une reconstruction 3D,

- ou bien, par l'estimation du mouvement de l'élément dans l'espace 3D, suivi d'une évaluation simultanée, basée sur les caractéristiques de l'élément, des projections de l'élément sur les plans image des caméras.

Ceci reste très complexe car la correspondance entre les informations 2D et 3D ne définissent pas une relation bijective. En effet, plusieurs configurations de l'élément dans l'espace 3D peuvent théoriquement correspondre aux caractéristiques de l'élément dans les images 2D. Les contraintes contenues dans les modèles cinématique et dynamique de l'élément peuvent être utilisées, entre autres, pour pallier ce problème et pour réduire la complexité en calcul des algorithmes [HUF04].

Acquisition du mouvement avec marqueurs

Des caractéristiques de l'élément peuvent être mises en avant, ou plus généralement ajoutées à l'élément, à l'aide de marqueurs qui seront plus simples à détecter et à identifier dans les images. L'utilisation de marqueurs demande un traitement supplémentaire dans la procédure qui consiste dans l'intégration de leur représentation numérique dans le modèle géométrique, car ils représentent des référentiels qui permettent de localiser l'élément. Avant l'acquisition, une manipulation supplémentaire de mise en place de marqueurs est également nécessaire (tâche qui est souvent considérée lourde et difficilement reproductible).

Leur utilisation repose sur le principe des « points éclairés en mouvement », initialement proposé par Johansson [Joh73]. La position des marqueurs est suivie dans le flux vidéo et leur trajectoire 3D est reconstruite par triangulation [Gle99]. L'infrarouge est l'une des techniques avec marqueurs les plus utilisées pour la capture optique de mouvements. Chaque marqueur émet généralement un rayonnement infrarouge distinct (entre eux et par rapport aux autres éléments de la scène), ce qui en fait des étiquettes très performantes pour repérer des éléments. En radiothérapie, ce type de marqueur a été d'un grand intérêt pour le positionnement et le contrôle en temps réel du patient par rapport à des points externes [BFOP00, SBR⁺06, TWW⁺06, PMF⁺07].

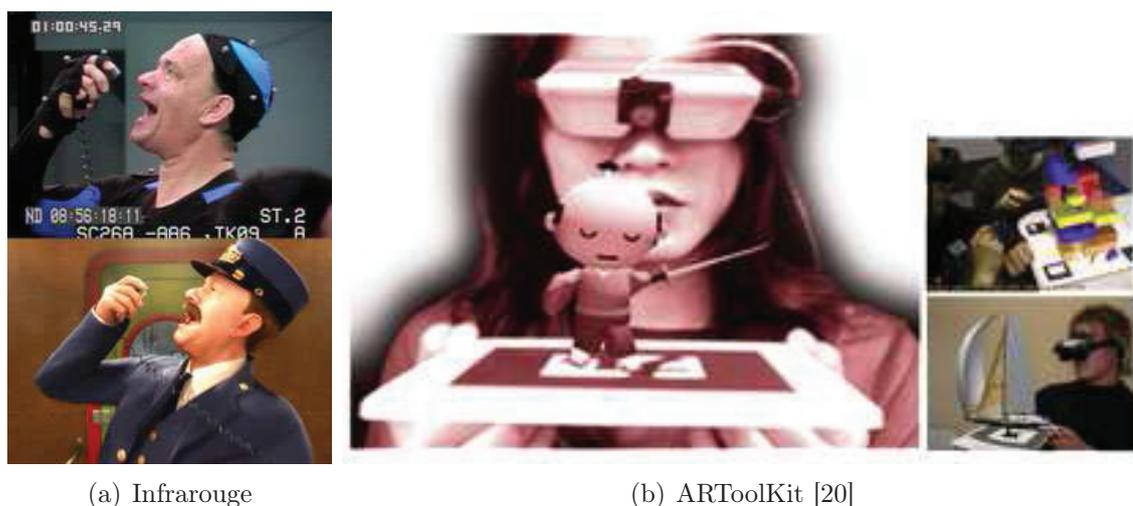


FIGURE 2.15: Exemple de système de suivi avec marqueur.

Les marqueurs de couleur sont un autre type de marqueurs utilisé couramment. Il s'agit souvent de gommettes (pastilles autocollantes) de différentes couleurs

en forme de disques (points) collées sur la surface de l'élément, bien que dans certains cas ces marqueurs puissent être dessinés directement sur la surface de l'élément [BKMM⁺04, CMMS05, BBA⁺07]. Analogue au rayonnement infrarouge, la couleur est utilisée pour identifier chaque marqueur.

Une variante des systèmes avec marqueurs est d'utiliser le recalage 3D [Bro92, ZF03] à la place de la reconstruction 3D. Ceci consiste à générer des vues synthétiques du marqueur à différentes positions et de les comparer à la vue réelle (image vidéo acquise) grâce à une mesure de distance. Dans la plupart des cas, cette technique permet d'obtenir des informations 3D à partir d'une seule caméra. La bibliothèque ARTToolkit [KBP⁺00] [20], ou sa variante ARTag [Fia05], mettent en place ce principe à l'aide de motifs codés. Dans le domaine médical, ces bibliothèques sont couramment utilisées pour leur simplicité [SAN⁺04, FNFB04, TMWG09, NPS⁺09].

Ces systèmes permettent d'atteindre généralement une précision submillimétrique et ils n'ont pas besoin d'une initialisation exhaustive, cependant l'une des principales difficultés algorithmiques est le suivi des marqueurs au cours du temps tout en évitant les ambiguïtés, ce qui met à défaut la robustesse du système. Une autre difficulté est la mise en correspondance des marqueurs dans toutes les vues [RL02]. Un post-traitement peut permettre de lisser les données et de résoudre les problèmes de correspondance, généralement corrigés « manuellement ».

Une autre variante des méthodes avec marqueurs, appelé lumière structurée, remplace les marqueurs physiques « collés » à l'élément par un motif projeté sur la surface de l'élément en utilisant des technologies comme le laser ou la projection d'images avec un vidéoprojecteur [MVL00, GDM05, SLB05, LMV⁺06, MLDR07]. La source lumineuse étant considérée comme l'une des caméras du système, et le motif étant connu au préalable, la mise en correspondance revient à calculer les correspondances entre les pixels « marqués » par le motif dans les images des caméras et les points du motif projeté [DW05]. C'est notamment la technique exploitée dans le système AlignRT®[17] [SMH⁺03], utilisé en radiothérapie pour le positionnement et le contrôle temps réel du patient par rapport à la surface externe (peau).



FIGURE 2.16: Exemple de système de suivi avec lumière structurée [NKY08].

L'une des principales difficultés de la technique de lumière structurée est que les points du motif ne sont pas associés à des points fixes dans l'espace, et donc un point du motif permet potentiellement de retrouver, dans deux images différents

d'une séquence, deux points différents dans la scène 3D. La mise en correspondance lors d'une acquisition peut être améliorée à l'aide d'une stratégie de codage du motif à projeter [SPB04, PSCF05].

Acquisition du mouvement sans marqueur

Lorsqu'on n'utilise pas de marqueurs, le problème de l'acquisition du mouvement devient plus complexe. Les techniques les plus courantes se basent sur l'extraction, dans les images, de primitives qui correspondent à des caractéristiques de l'élément à suivre. Ceci demande la mise en place de méthodes d'extraction de primitives, comme l'extraction de silhouettes par segmentation du fond qui consiste à calculer la différence entre une image d'une séquence vidéo, et un modèle de base de la même séquence vidéo [MGB05]. La détection des contours [HS88, DC01, FAHF08], la segmentation de régions [OKS80, GJH09], ou l'extraction de points caractéristiques [ST94], sont d'autres exemples de méthodes qui peuvent être utilisées.

Les différentes approches d'extraction de primitives ne garantissent pas l'ordre dans lequel elles sont extraites. La cohérence temporelle des primitives, qui permet de faciliter la corrélation d'informations, doit donc être assurée par un post-traitement des données. Il y a notamment trois cas à prendre en compte pour une primitive dans l'image courante par rapport à l'image précédente :

- standard : lorsque la correspondance entre deux primitives (une pour chaque image) est unique,
- fusion : lorsqu'une primitive dans une image correspond à plusieurs primitives dans l'autre image (par exemple, lorsque deux éléments ayant le même type de primitive se rapprochent, ou des effets produits par la perspective sur un même élément ayant plusieurs primitives du même type),
- disparition : lors qu'une primitive dans une image ne correspond à aucune primitive dans l'autre image (par exemple, en cas d'occultations des primitives ou lors qu'elles sortent du champ de vision).

Il existe également de méthodes qui exploitent le principe temporel afin de retrouver les primitives dans une image à partir des informations obtenues dans l'image précédente dans la séquence d'images. La méthode du flux optique est une méthode courante qui dérive d'un modèle de la vision humaine basé sur les mouvements apparents des objets, des surfaces, et des arêtes d'une scène visuelle provoqués par le mouvement relatif entre un observateur (un œil ou une caméra) et la scène [HS80, LK81, Bou00]. Un autre exemple est la méthode de Camshitf (Continuously Adaptive Mean Shift) [Bra98]. Elle est basée sur la couleur et utilise une technique robuste non paramétrique qui explore les gradients de densité pour trouver le mode¹³ (maximums locaux) des probabilités de distributions, appelé algorithme de « Mean Shift ». Les méthodes basées sur les textures¹⁴ utilisent des techniques de recalage d'images pour le calcul du mouvement de l'élément dans l'image.

Un autre défi est la corrélation de primitives entre les images de différentes caméras (avec acquisition synchronisée). Les méthodes de recalage d'images, qui consistent à calculer une transformation géométrique permettant de passer d'une image à une autre, sont couramment utilisées pour répondre à cette problématique

13. mode : valeur correspondant à la densité de probabilité maximum.

14. Une texture est une caractérisation de l'apparence fine des objets 3D à base d'images.

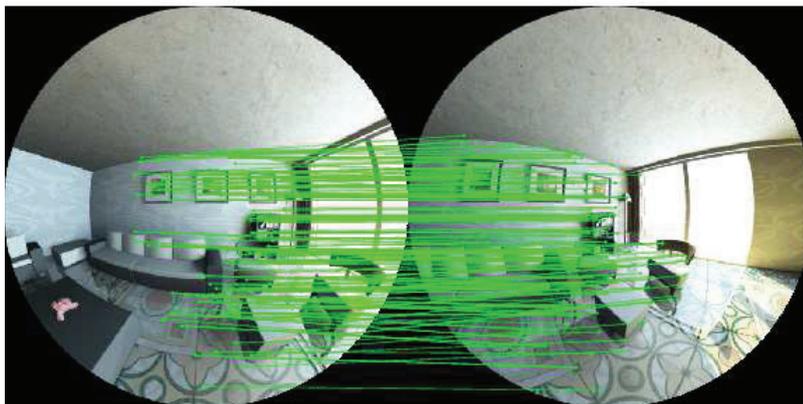


FIGURE 2.17: Exemple de corrélation de primitives entre deux images : utilisation de descripteurs SIFT sur des images omnidirectionnelles [AF10].

[Bro92, ZF03]. On retrouve également dans la littérature des méthodes d'extraction et corrélation, comme la méthode proposée par Lowe [Low04] basée sur l'analyse locale de l'image et le calcul de descripteurs SIFT (Scale-Invariant Feature Transform¹⁵).

Partiellement inspirée de cette dernière, la méthode SURF (Speeded Up Robust Features¹⁶) [BTVG06] se base sur des sommes de réponses d'ondelettes de Haar 2D et utilise efficacement les images intégrales. Lepetit et al. [LPF04] proposent une méthode de suivi temps réel par détection de primitives.

Des approches basées sur l'estimation du mouvement 3D exploitent davantage les informations contenues dans le modèle géométrique de l'élément (ou dans certains cas, de la partie à suivre de l'élément, par exemple, pour un humain, la tête, la main, les jambes, etc.). Le principe de base est centré sur l'équivalence de la projection de l'élément dans les plans image des caméras réelles et la projection du modèle géométrique dans les plans image des représentations numériques des caméras réelles. Pour calculer la meilleure approximation, l'ensemble des positions possibles, définies par les modèles cinématique et dynamique, est pondéré par rapport à l'équivalence des projections de l'élément en utilisant une mesure de distance. Ce procédé est connu sous le nom de recalage 3D, et les différentes méthodes existantes se différencient notamment par le choix des primitives et de la mesure de distance [HHC96, ST01, MCL01, DL05, PCMC06, SMB07, GJH09].

Au début du XXI siècle, des caméras 2.5D à temps de vol (en anglais, time of flight [TOF]) ont fait leur apparition. Ce système, composé d'une source de lumière infrarouge d'amplitude modulée et d'un capteur qui mesure l'intensité rétrodiffusée de la lumière infrarouge, permet le calcul de la distance entre les objets de la scène et la caméra [GYB04]. La carte de profondeur de la scène par rapport à la caméra est directement obtenue, mais ces systèmes ne proposent actuellement qu'une résolution assez faible, en tout cas insuffisante pour nos besoins. De plus, il est encore nécessaire de réaliser un post-traitement pour estimer la géométrie 3D de la scène, ce qui demande l'utilisation de plusieurs de ces capteurs [BG08].

15. en français : Transformation de Caractéristiques Visuelles Invariante à l'Échelle.

16. en français : Caractéristiques Robustes Accélérées.

2.4.3 Suivi du mouvement 3D de plusieurs éléments

La présence de plusieurs éléments augmente la complexité du suivi de mouvements. En effet, en plus des difficultés existantes pour le suivi d'un seul élément comme le changement d'apparence, le mouvement non rigide, la dynamique de l'éclairage et les occultations, on doit faire face à d'autres problématiques comme les similarités entre les éléments et leur priorités relatives.

Le domaine du suivi du mouvement 2D de plusieurs éléments dans une image a été largement exploré (par exemple, [Rei79, TSK00, HHD99, MB00, IM01, HCP01, ZN04, ORS09]), au contraire du suivi du mouvement 3D. Ce dernier, généralement réalisé par reconstruction à partir de plusieurs vues avec une précision faible, s'accompagne d'artefacts et de peu de prise en compte de l'orientation des éléments. L'utilisation de méthodes de suivi du mouvement d'un seul élément pour le suivi de plusieurs éléments est souvent peu envisageable pour des raisons de complexité algorithmique, qui augmente généralement proportionnellement au nombre d'éléments, car chaque objet est traité indépendamment.

[PLW08, WSB09] proposent des algorithmes de suivi temps réel pour plusieurs objets qui combinent le suivi basé sur des primitives par détection et par des corrélations image-à-image, la détection étant plus robuste mais nécessitant plus de temps de calcul. Cependant, la complexité en temps de calcul et en mémoire de ces algorithmes impose des limites dans le nombre d'éléments pouvant être traités simultanément. [PS10] a proposé récemment l'utilisation de techniques d'extraction de primitives dans les images, capables de traiter un grand nombre d'images simultanément [NS06], afin de résoudre les problèmes liés à l'adaptabilité aux nombre d'éléments, en assurant le traitement temps réel et la précision.

Les méthodes de suivi de mouvement de plusieurs éléments utilisant un modèle numérique géométrique donnent de meilleurs résultats en termes de précision [GD96, GVPG03, BGS07, SRB⁺07, KLW10]. Cependant, les méthodes proposées ont des complexités en temps de calcul plus importantes (que les méthodes n'en utilisant pas) et s'appuient sur des modèles simples comme des boîtes englobantes ou des modèles approchés composés de formes géométriques de base (cubes, cylindre, sphères, etc.).

2.5 Vidéosurveillance

Le premier système de surveillance par vidéo a été mis en place en 1942 par l'ingénieur allemand Walter Bruch pour le contrôle du lancement des fusées. Dans un effort pour combattre le crime, c'est dans les années 1970 que la vidéosurveillance devient très populaire, notamment en Angleterre afin de prévenir les attaques de l'Armée Républicaine Irlandaise. Elle est devenue ensuite un moyen très courant de persuasion et d'enregistrement de toute activité criminelle, utilisée de plus en plus dans les services publics (police, transports, administrations, etc.) et les entreprises privées (banques, magasins, etc.). En plus des applications en sécurité, de nombreuses autres applications de la vidéosurveillance ont vu le jour, comme la mesure du trafic routier, l'analyse des foules, le comptage démographique, etc. [MBCM97, SS03, HTWM04, YLPL05].

Depuis son origine, on a pu voir le développement de caméras plus performantes, pouvant même se déplacer ou changer d'orientation, voire être embarquées dans

des systèmes mobiles. Les techniques de visualisation et d'archivage d'images ont également été automatisées, pour maximiser le nombre des caméras à contrôler par un nombre de personnes donnée, et satisfaire le besoin croissant de surveillance 24h/24h.

2.5.1 Vidéosurveillance automatisée

La vidéosurveillance automatisée est aujourd'hui un domaine de recherche très dynamique qui veut exploiter l'aspect actif de la vidéo (moyen d'acquisition temps-réel) avec le but ultime de pouvoir « arrêter l'individu avant qu'il ne commette la faute ». L'utilisation de techniques de vision par ordinateur et de reconnaissance de formes, comme le suivi de mouvement ou la classification d'objets, permet l'acquisition de données sur les éléments de la scène, mais ceci demande la formalisation du contenu et de la structure de l'environnement afin que les informations puissent être traitées.



FIGURE 2.18: Exemple de suivi automatique de plusieurs objets dans un système de vidéosurveillance pour une zone résidentielle [YLPL05].

Citons aussi le besoin de pouvoir analyser ces données pour extraire de nouvelles informations afin d'interagir avec l'utilisateur du système, comme par exemple générer des alertes quand il est nécessaire que le personnel intervienne, pouvant faire appel à des techniques d'intelligence artificielle. Les différentes approches de la littérature pour résoudre cette problématique sont basées sur le principe que les actions qu'un « élément » réalise dans un ordre précis, et ses interactions avec d'autres éléments, peuvent être discriminantes pour la reconnaissance de résultats non acceptés ou non souhaités. Par exemple, lorsqu'on s'intéresse à la surveillance d'humains, la manière dont une personne agit n'est pas la même selon qu'il va commettre un acte répréhensible ou non. Ainsi, Chowdhury [CC03], qui s'intéresse à la détection d'événements dans un aéroport, propose une méthode de modélisation par la forme des trajectoires des éléments. Plusieurs solutions sont possibles en combinant les différentes techniques d'acquisition de données et d'analyse, en général sur deux types approches : l'analyse en fonction de la ressemblance totale ou partielle avec des données du passé (*a priori* ou par apprentissage), comme le Raisonnement à Partir de Cas (RàPC) [Kol92], ou l'analyse avec des modèles décrivant les évolutions de l'environnement, comme les Réseaux de Petri (RdP) [DA94] ou les chroniques [DG94].

Les états et les configurations de l'environnement sont analysés afin de dicter les futures actions et interactions des éléments et/ou de prendre des décisions qui vont influencer sur ces actions et interactions afin qu'elles suivent les modèles définis. Hongeng et al [HBN00b, HBN00a, HN01] utilisent des réseaux bayésiens et des Modèles

de Markov Cachés (MMC, ou HMM pour Hidden Markov Model) pour la reconnaissance de scénarios avec une seule caméra. Pour la reconnaissance d'événements, Hamid [HHE03] utilise des réseaux bayésiens dynamiques. Yamato [YOI92] propose un algorithme de reconnaissance du comportement humain, basé sur des MMC, et qui utilise des vecteurs pour caractériser des séquences d'images. D'autres travaux [MG01, ADM03, DT03] proposent également des méthodes basées sur des outils probabilistes pour la reconnaissance de scénarios.

2.5.2 Surveillance du patient dans le domaine médical

Dans le domaine médical, un nombre croissant d'appareils est utilisé pour la surveillance et le traitement thérapeutique des patients. Dans beaucoup de cas, la surveillance est justifiée par l'état de santé du patient ou par la complexité du traitement, le patient ayant besoin d'être suivi avec un soin tout particulier. L'un des principaux objectifs de la surveillance automatisée du patient est de détecter et d'identifier les situations acceptables ou interdites par rapport à sa santé. Cela est très informatif pour le personnel soignant, et doit être basé sur une analyse haut niveau des informations issues des différents outils de contrôle. Lorsqu'un protocole a été défini pour le traitement, cette surveillance automatisée se doit d'attester du bon suivi du protocole par le personnel médical. Mais cette tâche d'automatisation est difficile notamment parce que chaque patient est unique et son comportement est très aléatoire, et parce que les décisions médicales peuvent être basées sur des informations non accessibles aux outils de contrôle.

2.6 Réalité Augmentée

La Réalité Augmentée (RA) rend possible la superposition d'un modèle virtuel à ce que nous voyons naturellement de façon réaliste, afin de compléter notre perception du monde réel. Généralement, des images acquises à partir d'une caméra sont augmentées avec des objets virtuels 2D ou 3D. Des informations sur la scène filmée sont nécessaires pour donner l'impression que l'objet virtuel ajouté est dans une position réaliste et conforme aux autres objets de la scène (mise à l'échelle, effets de la gravité, etc.). Comme pour les Environnements Virtuels (jeux vidéo et autres), une des principales applications de la réalité augmentée est de permettre à l'utilisateur d'interagir avec des objets virtuels. Mais l'analyse de l'interaction entre les environnements réel et virtuel est également une grande source d'informations [DNT01]. Nous nous focalisons sur utilisation de la réalité augmentée dans le domaine médical ; un état de l'art général de la réalité augmentée peut être trouvé dans [Azu97, vKP10].

2.6.1 Réalité Augmentée dans le domaine médical

La principale motivation de l'utilisation de la RA dans le domaine médical est née du besoin de visualiser les données médicales et le patient dans un même espace physique. Il s'agit généralement de mettre en place un système en temps réel capable de fusionner les données enregistrées de l'intérieur du patient avec une vue de



(a) Un des premiers systèmes de RA [Sut65, Sut68]

(b) Application RA sur téléphone mobile – Wikitude Augmented Reality iPhone app. Photo : Karl Hilzinger – 2010

FIGURE 2.19: Réalité augmentée.

l'extérieur du patient afin d'assister le médecin ou le chirurgien. Une autre approche est la fusion des informations internes et externes acquises de manière synchronisée et en temps réel. Bajura et al. [BFO92] sont parmi les premiers à présenter ce type de système, en utilisant des images issues d'un examen d'ultrasons, dont les informations étaient intégrées à des images vidéo.

Dans la salle d'opération, chaque dispositif d'image introduit un nouvel écran, mais la fusion des images sur un même écran permet d'avoir la meilleure vue disponible. Aujourd'hui, de nombreux systèmes proposent également une visualisation 3D par stéréoscopie, facilitant l'interaction par rapport aux interfaces 2D [SFN08]. La RA offre donc une interface homme-machine plus intuitive, pouvant améliorer les performances de l'utilisateur. Par exemple, Hanna et al. [HSC98] ont montré que la position de l'affichage d'un endoscope a un impact significatif sur les performances du chirurgien.



(a) Superposition des structures d'intérêt pour la craniotomie [GEW+96]

(b) Guidage des opérations du foie [NPS+09]

FIGURE 2.20: Exemples de systèmes de RA dans le domaine médical.

Pour la mise en place de systèmes de RA, il faut faire appel à des méthodes qui sont encore aujourd'hui des sujets de recherche : calibrage et synchronisation des systèmes d'acquisition et d'interaction, recalage et fusion d'images multimodales, suivi de mouvement, simulation, etc..

Chapitre 3

Représentation de l'environnement

Notre solution pour la surveillance des séances de traitement en radiothérapie par vidéo est basée sur l'exploitation des informations déjà contenues dans le dossier médical du patient, ainsi que dans les procédures et modes opératoires de l'établissement de soins, et la vérification à tout instant de leur conformité par rapport à ce qui est vu par le système de caméras. L'un des premiers besoins qui se sont présentés a été la représentation numérique de ces informations afin qu'elles puissent être comparées aux informations extraites des images issues du système de caméras vidéo. Dans ce chapitre, nous commençons par décrire rapidement l'aspect global de la solution proposée, suivi de la présentation du système d'acquisition, du modèle générique pour la représentation des informations et de l'interaction entre les informations réelles et virtuelles.

3.1 Approche générale

L'idée générale de la solution proposée est de permettre à deux environnements d'interagir :

l'Environnement Réel (ER), c'est-à-dire la scène filmée par les caméras,

l'Environnement de Réalité Virtuelle (ERV) construit en utilisant un modèle numérique générique et contenant une description personnalisée de la scène : humains et objets (mobiles ou non).

L'interaction entre l'ER et l'ERV sera possible grâce au module de contrôle externe, qui est composé de sept sous-modules effectuant des tâches spécifiques pour le traitement des informations issues de deux environnements (les connections entre les sous-modules sont illustrées sur la Fig. 3.1).

Le module d'Initialisation du Modèle effectue les tâches liées à l'importation des informations *a priori* dont on dispose sur la scène, ainsi que la création de l'ERV (cf. section 3.3).

Le module d'Acquisition Vidéo est chargé de la gestion du système de caméras, de l'acquisition des images synchronisées et du traitement des images, afin qu'elles puissent être exploitées par les autres sous-modules (cf. section 3.2).

Le module de Suivi de Mouvement réalise notamment le calcul des positions des éléments à tout instant (cf. chapitre 4).

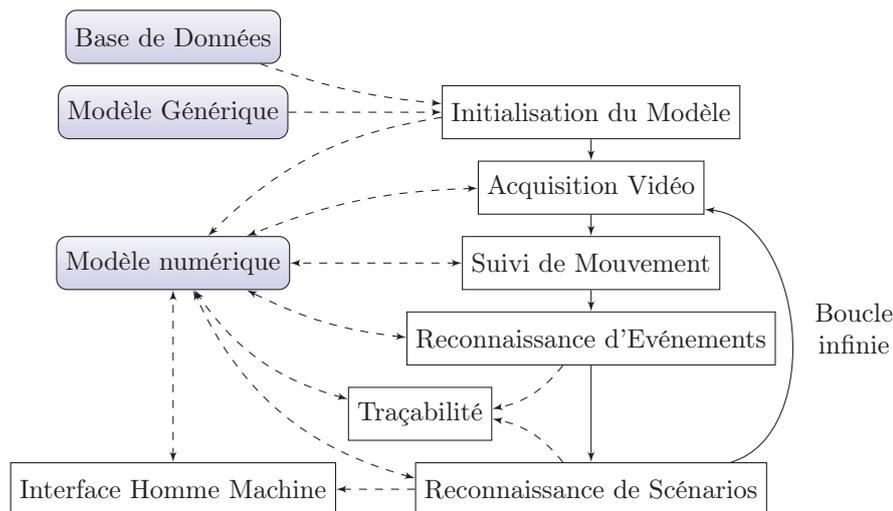


FIGURE 3.1: Schéma global de la solution proposée. Les lignes continues indiquent l'ordre dans le processus global et les lignes en pointillés indiquent le transfert de données.

Le module de Reconnaissance d'Événements inclut l'analyse des informations extraites à partir des images, afin d'obtenir de nouvelles informations (cf. section 5.3).

Le module de Reconnaissance de Scénarios accomplit les tâches de traitement des événements reconnus afin de détecter et d'identifier des scénarios sur une base de scénarios prédéfinis (cf. section 5.4).

Le module de Traçabilité assure l'enregistrement des données acquises pour définir un historique de chaque séance, et donc de tout le traitement d'un patient (cf. section 6.1).

Le module d'Interface Homme Machine permet l'interaction entre l'utilisateur et le module de contrôle externe (cf. section 6.2).

3.2 Système d'acquisition

Nous définissons le système d'acquisition comme un système composé d'au moins deux caméras vidéo. Les paramètres de chaque caméra sont exploités pour le traitement des images acquises dans le contexte 3D de la scène.

3.2.1 Calibrage du système

Le calibrage du système consiste dans le calcul, pour chacune des caméras, de la relation entre un point 3D dans la scène et le point 2D qui correspond à la projection du point 3D dans l'image. Ce calibrage se fait en trois étapes :

1. calibrage intrinsèque de chaque caméra,
2. calibrage extrinsèque de l'ensemble des caméras,
3. calibrage du système par rapport au repère global de la scène.

Modélisation des caméras

Nous utilisons le modèle sténopé (*pin-hole* en anglais), formulé par Abdel-Aziz et Karara [AAK71] et très utilisé dans la littérature, pour la représentation de chaque caméra du système. Le principe de ce modèle (sans distorsion) est qu'un point M (3D) dans la scène et le point m (2D) correspondant dans l'image se trouvent sur un rayon optique passant par le centre optique de la caméra. Comme on peut le voir sur la Fig. 3.2, quatre repères sont considérés dans ce modèle : le repère de la scène, R_w , le repère de la caméra, R_c , le repère du capteur, R_f , et le repère image, R_i . La relation entre les points m et M est donnée par trois transformations mathématiques élémentaires, chacune permettant de passer d'un repère à un autre. Cette relation est définie par l'équation suivante :

$$m = T_{if} T_{fc} T_{cw} M \quad (3.1)$$

où T_{if} est la transformation qui permet de passer de R_f à R_i , T_{fc} permet de passer de R_c à R_f et T_{cw} permet de passer de R_w à R_c .

Soit $(X, Y, Z, 1)$ les coordonnées homogènes de M dans R_w et (s_u, s_v, s) les coordonnées de m dans R_i , la représentation matricielle de Eq. 3.1 est donnée par :

$$\vec{m} = M_{T_{if}} M_{T_{fc}} M_{T_{cw}} \vec{M}$$

$$\begin{pmatrix} s_u \\ s_v \\ s \end{pmatrix} = \begin{bmatrix} k_u & s_{uv} & c_u \\ 0 & k_v & c_v \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} & \begin{matrix} t_x \\ t_y \\ t_z \end{matrix} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.2)$$

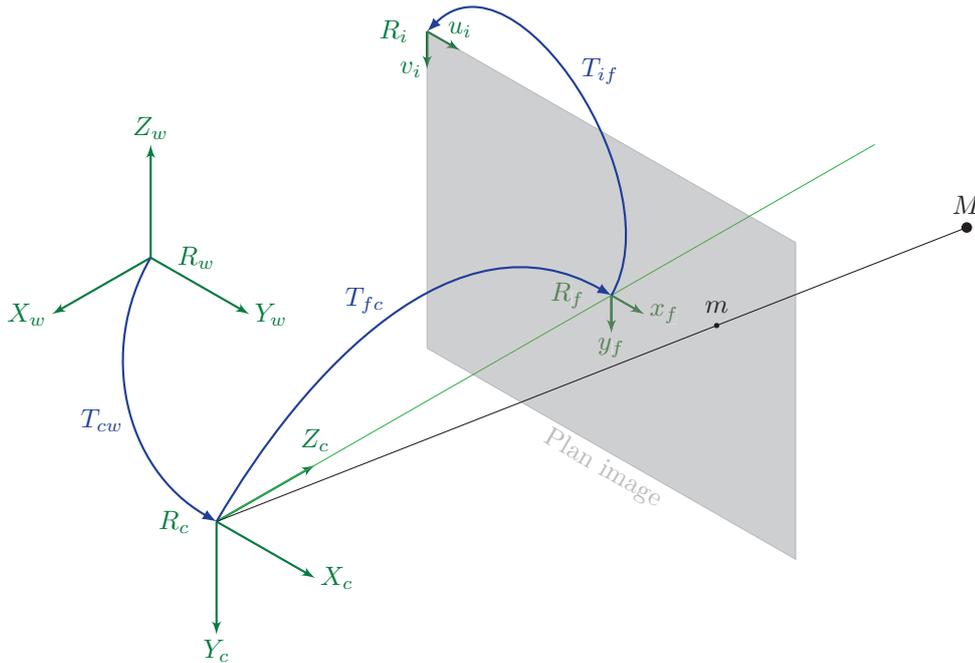


FIGURE 3.2: Modèle sténopé : transformations élémentaires et repères associés.

Le calibrage d'une caméra consiste à estimer deux types de paramètres contenus dans l'Eq. 3.2 :

les paramètres intrinsèques modélisent la géométrie interne et les caractéristiques optiques du capteur (transformations T_{if} et T_{fc}), dont :

- la distance focale f ,
- les facteurs d'agrandissement de l'image k_u et k_v ,
- les coordonnées de la projection du centre optique de la caméra sur le plan image (c_u, c_v) ,
- la non orthogonalité éventuelle des lignes et colonnes de l'image traduite par le facteur s_{uv} (généralement considéré de valeur nulle).

les paramètres extrinsèques représentent la relation entre le repère de la caméra et le repère de calibrage (transformation T_{fw}), dont :

- la localisation, donnée par le vecteur de translation (t_x, t_y, t_z) ,
- l'orientation, donnée par les matrices des rotations sur les trois axes $R_{3 \times 3}$ (angles d'Euler).

Lors de l'estimation des paramètres extrinsèques, le fait d'utiliser le même repère R_w pour le calibrage de toutes les caméras permet d'obtenir directement la relation entre les repères de chaque couple de caméras. Cette relation est définie par une multiplication matricielle des paramètres extrinsèques des caméras.

Calibrage dans le repère de la scène

Comme nous l'avons présenté à la section 2.2.2, les informations liées au traitement sont définies par rapport au repère à l'isocentre R_{iso} , qui est un repère fixe dans la salle de traitement. Il est donc nécessaire d'obtenir des relations entre les informations contenues dans le plan de traitement et les informations extraites des images. Pour cela les relations calculées pour le modèle des caméras sont utilisées, mais il faut aussi calculer la relation entre le repère R_{iso} et R_w (repère de calibrage utilisé pour l'estimation des paramètres extrinsèques des caméras du système). Cette relation correspond à une transformation T_{wiso} , ayant la même représentation matricielle que T_{fw} , c'est-à-dire une matrice 4×4 avec trois paramètres de localisation et trois paramètres d'orientation.

Le point M , défini dans R_w , sera dénoté M_{iso} dans R_{iso} . La relation entre M et M_{iso} est définie par l'équation suivante :

$$M = T_{wiso} M_{iso} \quad (3.3)$$

En combinant les équations Eq. 3.1 et Eq. 3.3, on obtient la relation entre le point M_{iso} et sa projection sur le plan image, le point m . Cette relation est donnée par l'équation :

$$m = T_{if} T_{fc} T_{cw} T_{wiso} M_{iso} \quad (3.4)$$

L'approche la plus courante est d'utiliser un dispositif définissant la relation T_{fw} comme une relation fixe et bien définie, et contenant deux types de repères visuels :

1. des repères pour le positionnement du dispositif à l'aide des lasers présents dans la salle, avec une relation fixe bien définie par rapport à R_{iso} ,



(a) Dispositif utilisé par [TWW⁺06] (b) Dispositif utilisé par [TMWG09]

FIGURE 3.3: Exemples de dispositif de calibration des caméras tenant compte de l'isocentre.

2. des repères permettant l'estimation des paramètres extrinsèques des caméras, avec une relation fixe bien définie par rapport à R_w .

La Fig. 3.3 montre deux exemples de dispositifs utilisés pour cette tâche de calibration. En raison des repères visuels que nous utilisons pour l'estimation des paramètres de localisation et orientation d'un élément (détaillés dans le chapitre 4) le dispositif que nous utilisons pour le calibration du système de caméras sera présenté à la section 6.5.

3.2.2 Modèle numérique du système d'acquisition

Une représentation numérique du système d'acquisition, contenant les paramètres intrinsèques et extrinsèques des caméras, est également intégré dans l'ERV. À chaque caméra $Rcam_i$ est associé un modèle numérique $Vcam_i$, où i est l'indice associé à une caméra du système d'acquisition. En exploitant les paramètres, $Vcam_i$ permet d'obtenir une image virtuelle $Vimg_{(i,t)}$ équivalente à une image réelle $Rimg_{(i,t)}$ acquise depuis $Rcam_i$ à tout instant t (à condition que l'ER et l'ERV soient bien corrélés). Dans le module de contrôle, la discrétisation du temps est faite en fonction de la fréquence d'acquisition du système d'acquisition. L'acquisition synchronisée d'un ensemble d'images (une par caméra) définit donc le temps t , et $Vimg_{(i,t)}$ est la simulation de l'acquisition d'une image à l'instant t depuis $Rcam_i$.

3.3 Modèle générique

De nombreux travaux se sont centrés sur la modélisation numérique des individus et des objets réels. En médecine, par exemple, des modèles permettant la simulation numérique de comportements physiques ont été développés dans le but d'être intégrés à des simulateurs médicaux [CDA99, TDG⁺00, BZSR09, SDV⁺10]. Ces modèles numériques sont souvent le résultat de la segmentation, manuelle ou assistée, des données issues de l'imagerie. Ils ont été conçus pour effectuer une tâche spécifique, et n'ont pas de caractère adaptatif ni générique. Dans le domaine du suivi du mouvement par vidéo, des travaux ont été présentés pour la modélisation des éléments de la scène [FS06, vdHDT⁺07, BMC08, PRD09, vdHHWD09, Sim10]. Ces travaux permettent la création de modèles de manière interactive, et peuvent

donc s'adapter à plusieurs configurations d'éléments, mais ils restent focalisés sur les caractéristiques visuelles.

Les traitements par radiothérapie sont des traitements très contrôlés, la radiothérapie étant la première discipline à avoir mis en place des outils de contrôle de qualité médicale [Ros02]. De ce fait, une grande quantité d'informations est déjà contenue dans le dossier du patient, notamment au niveau du plan de traitement. Ces informations sont définies dans des formats et repères établis pour la radiothérapie, principalement au format DICOM-RT¹. L'intérêt d'un modèle numérique de la scène est de pouvoir exploiter ces informations dans le cadre visuel de la surveillance en radiothérapie.

Un modèle générique permet la fusion des informations de différentes natures (physiques, géométriques, etc.), principalement pour enrichir les informations déjà présentes avec des informations qui autorisent, aident et/ou améliorent le contrôle des séances à l'aide de flux vidéo. La généricité recherchée dans la conception du modèle l'est également au niveau des instances des éléments présents dans la salle ; autrement dit, le modèle générique doit être capable de fournir un modèle numérique personnalisé aussi précis que possible pour tout type d'élément.

Les salles de traitement contiennent des machines de grande taille susceptibles d'en cacher entièrement ou partiellement d'autres, voire le patient par rapport à une caméra. Plus on dispose d'informations *a priori* sur la scène, plus on est capable de lever les ambiguïtés liées au système d'acquisition (occultation totale ou partielle, variation ou défaut d'éclairage, etc.). La modélisation de tous les éléments de la scène présents permet de mieux suivre les déplacements d'un élément donné grâce aux parties visibles de l'élément et aux informations des éléments qui l'entourent.

3.3.1 Description de la structure

Nous avons développé un modèle générique [PDM11] répondant à la problématique *M.1* présenté à la section 1.3.1. Ce modèle s'appuie sur les principes de la Géométrie de Construction de Solides (GSC, ou Constructive Solid Geometry [CSG]), qui permet de représenter un élément comme une combinaison de primitives géométriques élémentaires (par exemple, parallélépipèdes, cylindres, ellipsoïdes, etc.) à l'aide d'opérateurs géométriques booléens (par exemple, union, intersection, soustraction, etc.). Ce type de structure est bien adapté à la représentation d'éléments décomposables et permet la sauvegarde d'informations à l'aide d'un langage à balises, comme le langage XML².

Un élément est décrit dans le modèle par ses paramètres :

- visuels : couleur, texture, etc.,
- géométriques : dimension, forme, etc.,
- mécaniques : degrés de liberté, comportement des matériaux, etc.,
- sémantiques : informations diverses et exploitables sur l'élément comme les positions « cibles », le comportement de l'élément (de quels facteurs dépend sa présence ? se déplace t-il uniquement en présence d'un autre élément ? ...), etc..

1. DICOM-RT : Digital Imaging and COmmunications in Medicine - Radiation Therapy.

2. Langage de Balisage Extensible. XML pour eXtensible Markup Language.

Les éléments de la scène, ainsi que les différents composants de chaque élément, sont associés à l'aide d'opérateurs pouvant décrire des interactions physiques, géométriques et/ou sémantiques. On définit deux types d'éléments :

élément simple : pouvant être représenté par une primitive géométrique élémentaire, ou bien, un élément qui n'a pas besoin d'être subdivisé en fonction de ses paramètres mécaniques (par exemple, un élément sans articulation représenté par un maillage de forme quelconque),

élément complexe : composé d'éléments simples et/ou complexes.

La représentation par subdivision permet la réutilisation de certaines informations, comme celles liées aux caractéristiques présentes dans plusieurs éléments ou celles liées aux éléments toujours présents dans la salle. On peut donc considérer la création d'une bibliothèque d'éléments avec lesquels l'utilisateur est en mesure de définir n'importe quel élément, en ajoutant au fur à mesure des informations spécifiques à l'élément en cours de modélisation.

3.3.2 Intégration des informations dans le modèle

Nous introduisons la notion généralisée de *paramètre*, composé de :

- un identifiant unique (au sein d'un élément),
- un ensemble de couples « identifiant – valeurs », dont l'identifiant doit être unique au sein du paramètre et les valeurs seront alphanumériques,
- un ensemble de *catégories* ; une catégorie étant définie par un couple « identifiant – paramètre », dont l'identifiant doit être unique au sein du paramètre et le paramètre représente un sous-paramètre de l'élément (par exemple, la catégorie « couleur » contenant les valeurs de chaque canal RGB). Cela introduit une notion de récursivité dans la représentation des paramètres de l'élément.

Pour chaque élément, le seul paramètre obligatoire est le paramètre « identifiant », qui doit être unique pour chaque élément simple ou complexe dans la scène. Un élément contient également une matrice 4×4 qui représente la localisation (trois paramètres) et l'orientation (neuf paramètres) de l'élément dans la scène (Eq. 3.5). Cette matrice représente le référentiel de l'élément, et elle est définie dans un repère local si l'élément est contenu dans un élément complexe, ou dans le repère global de la scène sinon.

$$\begin{bmatrix} & & & t_x \\ & R_{3 \times 3} & & t_y \\ & & & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & t_x \\ 0 & 0 & 0 & t_y \\ 0 & 0 & 0 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} & & & 0 \\ & R_{3 \times 3} & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.5)$$

Il est important de signaler que les 9 paramètres liés à l'orientation peuvent être réduits à trois paramètres indépendants, un pour chaque rotation autour d'un axe. Par souci de simplicité de calcul (pour plus d'informations voir le chapitre 4), un vecteur de six paramètres représentant la pose $Pose3D_{j,t}$ de $Elem_j$ à l'instant t (la localisation : trois paramètres, et l'orientation : trois paramètres) est également contenu dans le modèle (Eq. 4.1). Les valeurs de ce vecteur et de la matrice sont dépendantes, car elles traduisent les mêmes informations.

3.3.3 Intégration du modèle dans le module du contrôle

Afin d'assurer une homogénéité dans les données, trois mesures principales ont été mises en place :

- un dictionnaire d'identifiants de paramètres, indiquant notamment les structures prédéfinies pour les catégories, la nature des données associées (par exemple : valeur numérique, chemin d'accès vers un fichier, identifiant d'un autre élément associé, etc.) et, dans certains cas, les valeurs possibles (par exemple, type d'élément : simple ou complexe, etc.),
- une bibliothèque de fonctions permettant la lecture et l'écriture de fichiers au format XML, que nous avons appelés « fichiers descripteurs », qui exploitent le dictionnaire d'identifiants. Ces fichiers contiennent les informations du modèle numérique d'une scène, ou bien d'une partie de la scène et/ou d'un de ses éléments,
- une bibliothèque de fonctions permettant l'importation d'informations issues de fichiers d'autres formats, qui implémentent une « traduction » d'informations à l'aide du dictionnaire d'identifiants.

Les différents algorithmes implémentés dans le module de contrôle font appel au dictionnaire d'identifiants afin de retrouver, ajouter et/ou mettre à jour les informations dans le modèle numérique.

3.3.4 Descripteurs 3D

Le point fort du modèle repose sur la notion de « descripteur ». Un descripteur est défini comme une caractéristique ou un objet associé à l'élément qui permet de le reconnaître visuellement (mais pas forcément de manière unique). On n'est pas forcément intéressé par l'élément tout entier, mais par ses caractéristiques particulières qui sont moins difficiles à extraire dans les images et qui permettent, après l'extraction de quelques-unes d'entre elles, de suivre les mouvements de l'élément grâce à leur définition dans le modèle numérique. Cette notion introduit le principe de multi-primitives, qui donne des résultats plus précis et robustes lors de la détection et le suivi de mouvement d'éléments dans des images vidéo [MC02].

Un descripteur sera souvent associé à une caractéristique déjà contenue dans la surface de l'élément, comme une région de couleur ou un motif, ou bien à un objet pouvant être fixé à la surface, comme des gommettes de couleur ou des balles à réflexion d'infrarouge. La notion de descripteur est donc implémentée comme étant la représentation numérique de la scène (espace 3D) qui, par projection dans le plan image des caméras en fonction des paramètres intrinsèques et extrinsèques, correspond à des primitives 2D (contours, points caractéristiques, motifs, textures, etc.). Soit $Desc_{(j,k)}$ le descripteur d'indice k contenu dans l'élément $Elem_j$. On définit $VSimg_{(i,t,j,k)}$ comme l'image virtuelle obtenue à partir de $Vcam_i$ à l'instant t et qui contient uniquement la projection de $Desc_{(j,k)}$. Si l'on considère qu'on dispose d'outils permettant d'extraire les différents type de primitives définies dans les descripteurs, on peut définir $RSimg_{(i,t,j,k)}$ comme le résultat de la segmentation de $Rimg_{(i,t)}$ et contenant uniquement la primitive associée à $Desc_{(j,k)}$. Si l'ER et l'ERV sont bien corrélés, on peut donc dire que $VSimg_{(i,t,j,k)}$ et $RSimg_{(i,t,j,k)}$ sont équivalentes. On définit enfin $Vprim_{(i,t,j,k)}$ comme l'ensemble des pixels non-nuls dans $VSimg_{(i,t,j,k)}$ (de même pour $Rprim_{(i,t,j,k)}$ et $RSimg_{(i,t,j,k)}$).

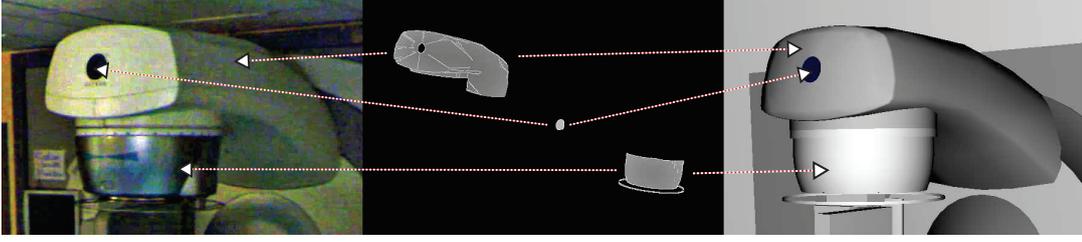


FIGURE 3.4: Exemple de descripteurs pouvant être utilisés (cas du bras d'irradiation de l'accélérateur linéaire du Centre Léon Bérard [7]) : deux descripteurs « région » et un descripteur « logo » (en forme de disque bleu) ont été établis par rapport à la couleur des surfaces de l'élément. À gauche : vue réelle de la scène ; au milieu : vue virtuelle du bras d'irradiation décomposé par rapport aux descripteurs. À droite : vue virtuelle de la scène.

Dans le modèle générique, un descripteur est un cas particulier d'un élément simple. Un élément utilisant des descripteurs est donc modélisé par un élément complexe contenant des descripteurs, et des éléments simples et/ou complexes. Les paramètres de la pose $Pose3D_{d,t}$ (défini à l'aide de l'Eq. 4.1 – où d est l'indice du descripteur en tant qu'élément simple, et donc $Desc_{(j,k)}$ équivaut à $Elem_d$) ont des valeurs fixes, c'est-à-dire, égales pour tout t . Pour rappel, ces paramètres sont définis dans le repère de l'élément qui les contient.

Les descripteurs représentent donc des référentiels pour l'élément auquel ils sont associés. De ce fait, si l'on est capable de calculer la transformation $T_{Desc_{(j,k)}iso}$ qui permet de passer du repère global de la scène R_{iso} au repère du descripteur $R_{Desc_{(j,k)}}$, la pose de l'élément par rapport à R_{iso} est également disponible car la transformation entre les repères de l'élément et du descripteur, que nous appellerons $T_{Pose3D_{d,t}}$, est une donnée connue. La transformation $T_{Pose3D_{j,t}}$ qui permet de passer du repère de l'élément $Elem_j$ au repère R_{iso} est donc définie par l'équation :

$$T_{Pose3D_{j,t}} = T_{Pose3D_{d,0}}^{-1} T_{Desc_{(j,k)}iso} \quad (3.6)$$

où $T_{Pose3D_{d,0}}^{-1}$ est la transformée inverse de $T_{Pose3D_{d,0}}$.

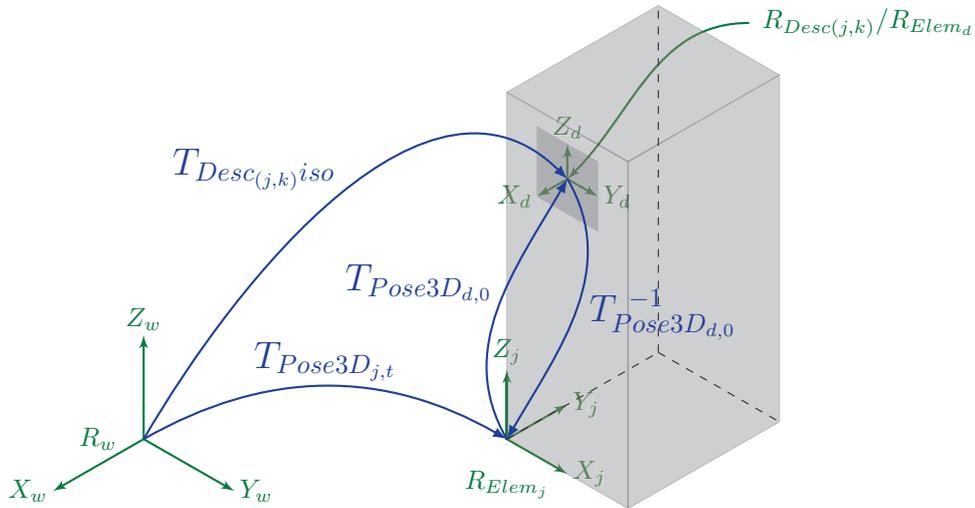


FIGURE 3.5: Transformations mathématiques entre les repères : scène, élément et descripteur.

Les descripteurs peuvent avoir d'autres utilisations dans le module de contrôle. Par exemple, il est possible de définir des zones d'intérêt (au niveau image) par rapport à un élément en générant des images virtuelles contenant la silhouette d'un descripteur associé au « volume minimum englobant » d'un élément. En utilisant cette zone d'intérêt, obtenue à l'instant $t - 1$, et les caractéristiques mécaniques de l'élément, il est possible de prédire la zone (au niveau image) dans lequel l'élément est potentiellement présent à l'instant t .

3.3.5 Opérateurs dans le modèle numérique

Le dernier concept dans le modèle numérique est celui d'« opérateur ». Un opérateur définit soit une action liée à un élément, soit une interaction entre plusieurs éléments. Les actions ou les interactions associées peuvent être d'ordre physique, géométrique et/ou sémantique. Par exemple, dans la Fig. 3.6 :

- l'opérateur « toujours connecté » définit une interaction physique entre les deux éléments, indiquant que pour n'importe quel déplacement leurs surfaces externes sont toujours en contact,
- des « référentiels locaux » peuvent également être traduits en opérateurs définissant une interaction géométrique, pour définir par exemple une distance minimale ou fixe entre deux éléments,
- l'opérateur « toujours au-dessus » met en place une notion sémantique, en indiquant que l'objet vert ne se déplace que sur le dessus de la surface de l'objet bleu (toujours en référence à la Fig. 3.6).

Une bibliothèque de fonctions contient les différentes implémentations des actions ou des interactions possibles. Un opérateur est donc composé d'un identifiant unique (au sein de l'élément), d'un identifiant permettant d'accéder à la fonction associée, et d'un ensemble de paramètres qui seront utilisés comme des données d'entrée pour la fonction associée (notamment contenant les identifiants des éléments associés).

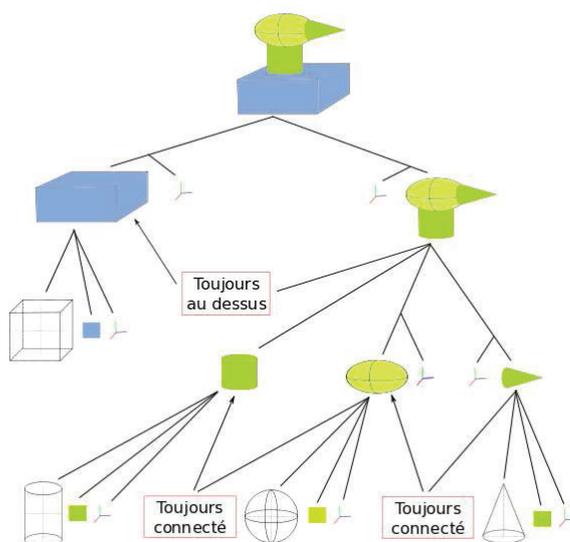


FIGURE 3.6: Exemple d'arborescence d'un modèle numérique issue du modèle générique.

3.3.6 Initialisation du modèle personnalisé

L'initialisation du module de contrôle commence par la mise en place et le calibrage du système d'acquisition (cf. section 3.2). À condition que les caméras ne soient pas déplacées, ces opérations ne seront pas répétées (bien que dans la pratique, des manipulations de maintenance sont nécessaires pour vérifier qu'il n'y a pas eu de changement dans les paramètres du système d'acquisition).

La prochaine étape est l'initialisation du modèle personnalisé de la scène, notamment pour les objets (machines et autres) toujours présents dans la salle de traitement. La seule contrainte est que le fichier descripteur associé à la salle soit bien défini, c'est-à-dire qu'il contienne bien une description de la salle à vide ainsi que tous les éléments potentiellement présents. Considérant que chaque élément a une configuration de départ, dite « position zéro », il suffit de placer physiquement tous les éléments dans cette position et de définir cette position pour tous les éléments dans le modèle numérique. Si cette tâche est bien réalisée, nous disposons de deux environnements, l'ER et l'ERV, bien corrélés.

Le modèle est ensuite réinitialisé à chaque séance, avec une conservation temporelle des informations des éléments encore présents dans la salle entre les deux séances successives (celle qui vient de se terminer et la prochaine) et une intégration des données du patient (notamment le plan de traitement) de la séance qui va commencer.

3.4 Environnement de Réalité Augmentée

La fusion de l'ER et l'ERV permet d'obtenir un environnement de Réalité Augmentée. Si les deux environnements sont bien corrélés, les images $Rimg_{(i,t)}$ et $Vimg_{(i,t)}$ le sont aussi (Fig. 3.7). Le contraire n'est pas forcément vrai, mais la corrélation d'images réelles et virtuelles permet de corrélérer certaines informations, principalement sur les éléments vus par les caméras. La corrélation entre les environnements peut être ensuite complétée, corrigée et/ou améliorée grâce aux connaissances *a priori* contenues dans le modèle. Il est intéressant de faire remarquer que la superposition des images $Rimg_{(i,t)}$ et $Vimg_{(i,t)}$ offre déjà un outil de contrôle visuel sur la corrélation des deux environnements [SFN08, TMWG09].

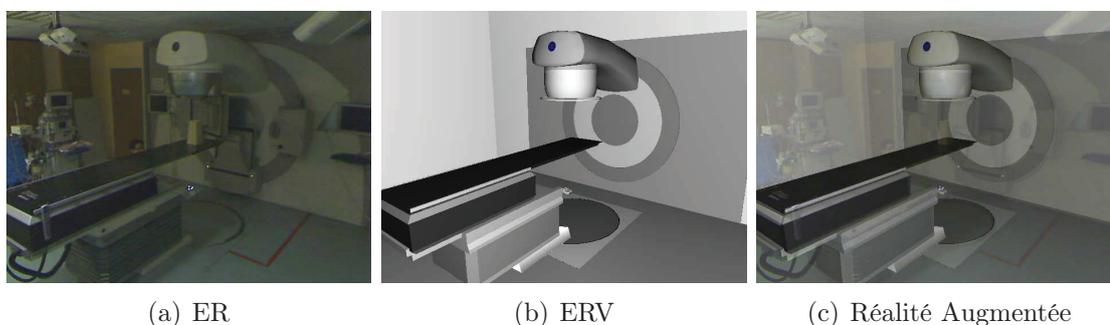


FIGURE 3.7: Réalité Augmentée à partir de la fusion des environnements ER et ERV.

Chapitre 4

Calcul de la pose

Nous disposons d'un environnement de réalité augmentée bien défini, dont les deux environnements réel et virtuel sont bien corrélés au début de l'exécution du module de contrôle par une procédure d'initialisation présentée dans le chapitre précédent. Pendant l'exécution du module, l'ERV est mis à jour en fonction de ce qui est vu par le système d'acquisition dans le l'ER, c'est-à-dire par l'exploitation des informations extraites des images vidéo. Dans ce chapitre nous présentons la méthode que nous avons implémentée pour le suivi de mouvement des éléments dans la scène, en commençant par une formalisation des besoins, puis une description détaillée de la méthode du Recalage 3D, et pour finir, la présentation des différents outils permettant de compléter et/ou d'améliorer la méthode par rapport aux problématiques en vision par ordinateur et en optimisation présentées à la section 1.3.1.

4.1 Pose 3D

Le terme « pose », couramment utilisé dans les domaines de vision par ordinateur et de robotique, correspond à l'ensemble de valeurs des paramètres donnant des informations sur la localisation et l'orientation d'un élément dans un repère bien défini. Dans un repère cartésien à trois dimensions, la pose d'un élément est représentée par un vecteur contenant six paramètres :

$$Pose3D = (Tx, Ty, Tz, Rx_0, Ry_1, Rz_2) \quad (4.1)$$

où les trois premiers paramètres représentent la localisation de l'élément, et constituent, dans le repère global de définition de la pose, les valeurs du vecteur translation $T(Tx, Ty, Tz)$ entre le centre O_g du repère global et le centre O_e du repère de l'élément. Les trois autres paramètres représentent l'orientation de l'élément et définissent des rotations selon les angles de Tait-Bryan dans l'ordre $X Y' Z''$, plus communément connus comme angles d'Euler [Gav99, DF01]. Dans ce contexte, les rotations sont réalisées de manière séquentielle, c'est-à-dire, qu'on réalise une succession de trois mouvements de rotation autour de trois axes perpendiculaires (chaque rotation sur un axe prédéfini). Chaque rotation est réalisée en s'appuyant sur le repère de l'élément après que la transformation antérieure a été réalisée, et les axes de rotation sont ensuite redéfinis. Nous avons fait ce choix de représentation parce que la relation entre les rotations et les degrés de liberté d'un élément est plus intuitive,

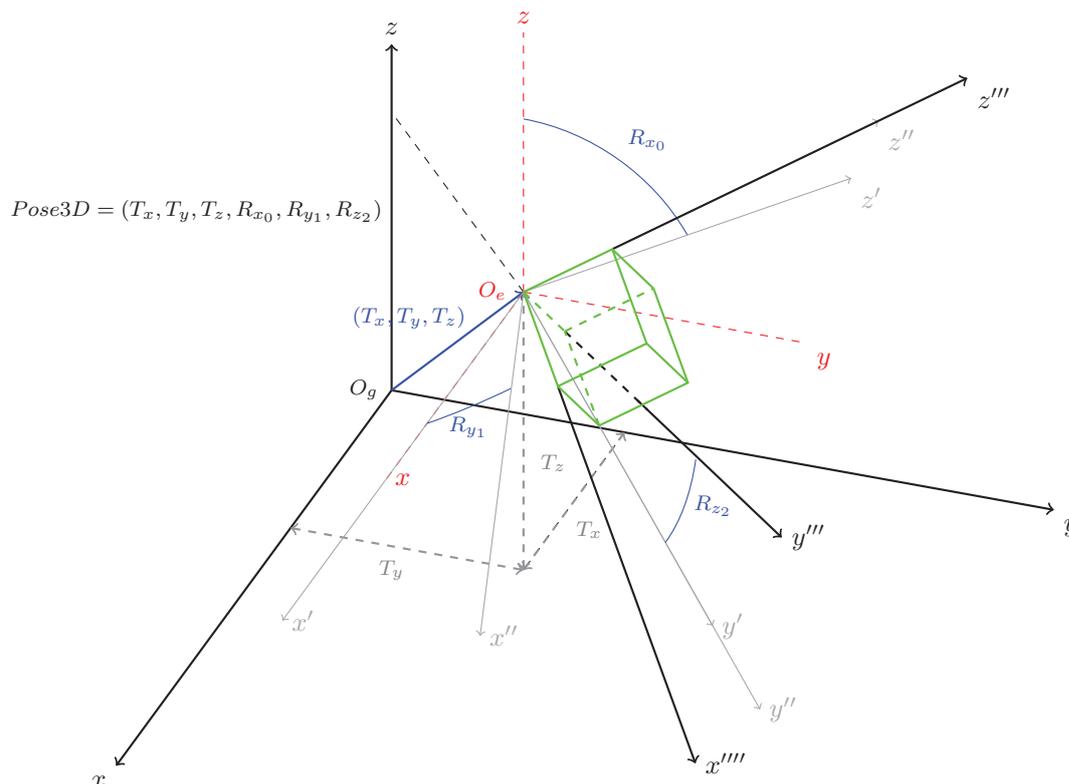


FIGURE 4.1: Pose 3D.

dû notamment au fait que les machines présentes dans la salle de traitement ne réalisent qu'une rotation sur un axe à la fois, mais aussi parce que la dynamique des éléments (probabilité que l'élément se déplace et amplitude des déplacements) dans la scène et la probabilité d'arriver à la « perte d'un degré de liberté »¹ ont des valeurs faibles.

La pose 3D peut être également représentée par une matrice 4×4 , contenant 12 paramètres dépendants des six paramètres du vecteur qui représente également la pose (Eq. 4.1) :

$$M_{Pose3D} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.2)$$

Cette matrice peut être également décomposée en deux (voir aussi Eq. 3.5) : une matrice translation et une matrice rotation. Cette dernière peut être décomposée encore en trois matrices, une pour chaque paramètre de rotation. On obtient donc :

$$M_{Pose3D} = M_T \times M_{R_{x_0}} \times M_{R_{y_1}} \times M_{R_{z_2}} \quad (4.3)$$

avec :

1. Le blocage de cardan (*gimbal lock* en anglais) survient, dans un espace à trois dimensions, lors que deux ou trois axes de rotations sont portés par la même direction (axes parallèles) engendrant la perte d'un degré de liberté.

$$M_T = \begin{bmatrix} 0 & 0 & 0 & T_x \\ 0 & 0 & 0 & T_y \\ 0 & 0 & 0 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad M_{R_{x_0}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(R_{x_0}) & -\sin(R_{x_0}) & 0 \\ 0 & \sin(R_{x_0}) & \cos(R_{x_0}) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{R_{y_1}} = \begin{bmatrix} \cos(R_{y_1}) & 0 & \sin(R_{y_1}) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(R_{y_1}) & 0 & \cos(R_{y_1}) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad M_{R_{z_2}} = \begin{bmatrix} \cos(R_{z_2}) & -\sin(R_{z_2}) & 0 & 0 \\ \sin(R_{z_2}) & \cos(R_{z_2}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

L'estimation de la localisation et de l'orientation d'un élément dans la scène sera réalisée par l'estimation des paramètres de translation et de rotation de la pose. Dans notre cas, les paramètres de tout élément sont définis par rapport à un repère global unique, confondu avec celui de l'isocentre.

4.2 Recalage

Le recalage est une technique ayant de nombreuses applications, qui consiste à déterminer la relation spatiale entre deux repères associés à un même élément. L'approche générale repose sur la recherche de la transformation mathématique $\hat{\Gamma}$ qui permet d'optimiser un critère de similarité entre deux représentations de l'élément : la représentation de l'élément dans l'un des repère, R_1 , et sa représentation dans le deuxième repère, R_2 , à laquelle la transformation $\hat{\Gamma}$ a été appliquée. Ceci peut être formulé par l'équation suivante :

$$\hat{\Gamma} = \underset{\Gamma \in E\Gamma}{\text{arg min}} C(E_{R_1}, E_{R_2} \circ \Gamma) \quad (4.4)$$

où $E\Gamma$ est l'espace de transformations possibles pour l'élément (espace de recherche), C est la mesure du critère de similarité (ou de dissimilarité), E_{R_n} est la représentation de l'élément dans le repère R_n (avec $n \in \{1, 2\}$), et $E_{R_2} \circ \Gamma$ représente l'application de la transformation Γ à la représentation E_{R_2} de l'élément.

Il existe deux approches majeures en fonction du type d'informations de l'élément qui seront exploitées dans le critère de similarité (bien que des méthodes mixtes aient également été proposées) :

les approches géométriques sont basées sur la mise en correspondance de primitives géométriques extraites dans les représentations de l'élément (points, normales, etc.). Le choix des primitives est fait en fonction d'un certain nombre de propriétés, comme la complexité en temps de calcul, la robustesse et la précision de leur détection [ZF03]. L'utilisation de primitives réduit la complexité en nombre de données à traiter, portant sur des informations de haut niveau. Cependant, le recalage n'est assuré qu'au niveau des primitives, et il est sensible aux erreurs de segmentation et d'extraction des primitives.

les approches denses (ou iconiques) se fondent sur la comparaison des valeurs (ou des vecteurs de valeurs) associés, dans l'espace discret, aux représentations de l'élément. L'exemple le plus courant est l'intensité du niveau des gris des pixels dans une image. Bien que ces approches ne demandent pas d'étape supplémentaire de réduction des données, elles ont besoin de la mise en place de stratégies d'interpolation des valeurs car les discrétisations de chacune des représentations de l'élément peuvent ne pas correspondre entre elles. Malgré la

complexité en nombre de données à traiter, ces approches ont l'avantage d'être complètement automatiques. De plus, des approches multirésolution peuvent réduire la complexité en nombre de calculs pour le traitement de ce type de données.

Une fonction permettant de mesurer la similarité, ou la différence, entre les représentations de l'élément est associée au critère de similarité. Cette fonction prend notamment en compte le type d'information choisie, et prend en entrée les représentations de l'élément. L'application des transformations sur les informations de l'élément doit suivre un modèle de déformation :

la transformation rigide suppose que l'élément est rigide (sans déformation), et uniquement les paramètres de translation et de rotation sont estimés,

la transformation métrique estime, en plus des paramètres de translation et de rotation, un facteur d'échelle isotrope (invariant à la direction),

la transformation affine estime, en plus des paramètres de translation et de rotation, un facteur d'échelle anisotrope et des paramètres de transvection (cisaillement),

la transformation projective estime les paramètres définissant une relation entre deux espaces projectifs (sous-espaces vectoriels),

la transformation non-linéaire estime les paramètres permettant de modifier localement et globalement la forme de l'élément.

Les différentes méthodes d'optimisation de la recherche de $\hat{\Gamma}$ dans $E\Gamma$, focalisées sur le terme minimisation (*arg min*) de Eq. 4.4, peuvent être classées en 4 principaux types :

les méthodes de résolution explicite sont celles qui acceptent une solution explicite. Ces méthodes sont peu nombreuses et se limitent principalement aux méthodes géométriques mettant en jeu des énergies quadratiques (par exemple, minimisation de la somme du carré des erreurs, etc.),

les méthodes d'ordre 0 utilisent l'évaluation du critère de similarité sur un certain nombre de caractéristiques de l'élément. La recherche exhaustive dans l' $E\Gamma$ est bien la méthode la plus simple permettant d'obtenir un résultat optimal, mais elle est trop coûteuse en temps de calcul. D'autres méthodes parcourent l' $E\Gamma$ de manière itérative en utilisant les informations des transformations déjà évaluées, dont les plus populaires sont la méthode du simplexe [NM65] et la méthode de Powell [Pow64]. Le principal inconvénient de ces méthodes itératives est qu'elles ne garantissent pas la convergence vers le minimum global,

les méthodes d'ordre supérieur utilisent des informations additionnelles sur la mesure du critère de similarité à optimiser. On retrouve les méthodes de descente de gradient et du gradient conjugué, qui exploitent le gradient de la fonction de la mesure, et les méthodes de Newton, qui exploitent le hessien² de la fonction de la mesure, ou encore l'algorithme présenté par Marquardt [Mar63] qui combine les méthodes basées sur le gradient et le hessien. Ces méthodes permettent d'augmenter la vitesse de convergence vers le minimum mais, comme les méthodes d'ordre 0, elles ne garantissent pas la convergence vers le minimum global,

2. discriminant d'une matrice hessienne d'une fonction numérique.

les méthodes non déterministes ou stochastiques utilisent des mécanismes de transition probabilistes et aléatoires, notamment pour accélérer le processus de recherche. Les méthodes les plus utilisées sont basées sur des algorithmes stochastiques d'optimisation itératifs, appelés métaheuristiques [BR03], dont beaucoup s'inspirent des comportements observables dans la nature. On retrouve trois principaux algorithmes d'optimisation non déterministes :

- les algorithmes évolutionnaires, basés sur la « théorie de l'évolution » (biologie), dont une des principales familles est celle des algorithmes génétiques [Gol89, WDM01] qui s'appuient sur des techniques dérivées de la génétique et de l'évolution naturelle,
- le recuit simulé [KGV83], dont le principe est inspiré du processus du recuit³ en métallurgie,
- la recherche « tabou » [Glo86], dont l'idée principale est de pouvoir mémoriser et exploiter l'historique de l'exploration de solutions.

L'utilisation d'une notion de hasard permet à ces méthodes de pallier les problèmes dus à des minima locaux, et éventuellement d'accéder à la solution optimale, bien que malheureusement leur vitesse de convergence soit faible et qu'elles ne garantissent pas que la solution sera trouvée.

4.2.1 Recalage 3D pour le suivi de mouvement

En vision par ordinateur et en robotique, le recalage 3D est une technique exploitant les informations contenues dans le modèle numérique d'un élément dans le but de suivre ses mouvements sur les images d'un flux vidéo. Elle est basée sur le calcul de la pose 3D de l'élément par un procédé itératif de comparaison entre images réelles et virtuelles, ces dernières contenant la projection du modèle numérique.

Plusieurs travaux, démontrant la faisabilité de la technique, ont été présentés pour le suivi de personnes ou de parties de leur corps, en utilisant des modèles humanoïdes (volumiques ou surfaciques) [OH99a, ST01, GJH09], mais aussi pour le suivi d'objets (avec des modèles rigides ou non), notamment appliqué à l'« asservissement visuel »⁴ [HHC96, PCMC06]. Les travaux basés sur le recalage 3D que l'on peut trouver dans la littérature, bien qu'ils utilisent des modèles personnalisés aux éléments cibles, ont comme principales différences le choix du critère de similarité et de la méthode d'optimisation.

Lorsqu'on dispose d'un modèle numérique de l'élément à suivre il est possible, comparativement à celles n'utilisant pas, d'assurer une meilleure robustesse face aux bruits générés par le système d'acquisition et le traitement des images vidéo, ainsi qu'une meilleure précision grâce aux projections de l'élément permettant d'avoir des informations plus fidèles sur l'apparence de ce dernier dans les images vidéo [VLF04, LF05]. Cependant, le principal inconvénient du recalage 3D est sa complexité en nombre de calculs, et son utilisation est donc rarement envisageable pour des applications temps réel ou interactives, sauf à se trouver dans des situations où les contraintes permettent de réduire considérablement les espaces de recherche, comme nous allons le voir.

3. le recuit d'une pièce métallique est un procédé correspondant à un cycle de chauffage.

4. l'asservissement visuel est une technique qui utilise des informations obtenues à partir de senseurs visuels afin de réaliser le contrôle de la position d'un robot (en anglais : *visual servoing*).

4.3 Formalisation de la solution proposée

Les caractéristiques propres aux salles de traitement en radiothérapie permettent d'envisager l'utilisation du recalage 3D pour le suivi robuste et précis du mouvement des éléments à la salle (patient, personnel, machines et autres objets – un exemple visuel d'une salle de traitement est donné dans Fig. 4.2). Ces caractéristiques sont dues principalement à la complexité du traitement (du point de vue médical et dosimétrique), et il est notamment possible d'exploiter :

- les informations sur le patient contenues dans le dossier du patient, et plus précisément dans le plan de traitement et dans les études d'imagerie réalisées en amont du traitement. Ces informations sont souvent très précises, surtout dans la zone à irradier,
- la définition de la scène comme milieu hautement contrôlé (protocoles et consignes de sécurité précises et bien définies, entre autres),
- la faible dynamique de la scène, notamment lors de l'irradiation car aucun déplacement non programmé n'est toléré (nous nous intéressons à la détection et la mesure des déplacements imprévus),
- le faible variabilité au niveau du contenu de la scène :
 - par rapport à deux séances consécutives, car c'est principalement l'élément « patient » qui sera différent, et une grande partie des éléments (machines et accessoires notamment) sont toujours présents dans la salle,
 - par rapport aux autres salles de traitement, car les modèles de machines sont peu nombreux (peu de constructeurs et coût élevé).

L'intégration de ces informations dans le modèle numérique, et l'exploitation de ce dernier, permet de réduire l'espace de recherche et d'optimiser la complexité en volume de calculs de la méthode de recalage 3D, comme nous allons le détailler dans la suite de chapitre. Remarquons que dans le modèle numérique les propriétés des éléments sont prises en compte, ainsi que les opérateurs définissant des interactions entre les éléments, qui vont également alléger le volume de calculs.



FIGURE 4.2: Exemple d'une salle de traitement en radiothérapie (Centre Léon Bérard [7]).

Nous nous intéressons au recalage rigide de la représentation virtuelle 3D d'un élément (modèle numérique) par rapport à ses représentations réelles 2D (projections sur le plan image des caméras du système). Le principe de base est le suivant : à un instant t , si la pose d'un élément de la scène dans l'ER correspond exactement à celle de sa représentation numérique dans l'ERV alors, pour toute caméra $Rcam_i$, la projection de l'élément sur le plan image dans l'ER est équivalente à la projection du modèle numérique de l'élément sur le plan image de $Vcam_i$ dans l'ERV.

Grâce à la faible dynamique de la scène, nous pouvons considérer que la pose calculée à l'instant $t - 1$ est une bonne approximation de la pose à l'instant t (Fig. 4.3). Nous cherchons donc à calculer un vecteur :

$$\Delta Pose3D_{(j,t)} = (\Delta T_x, \Delta T_y, \Delta T_z, \Delta R_{x_0}, \Delta R_{y_1}, \Delta R_{z_2})$$

tel que :

$$Pose3D_{(j,t)} = Pose3D_{(j,t-1)} + \Delta Pose3D_{(j,t)} \quad (4.5)$$

Nous exploitons la notion de descripteurs, définie à la section 3.3.4, pour mettre à jour le modèle numérique de l'élément par rapport à ses représentations réelles dans les images vidéo. Cette notion inclut l'aspect multi-primitives, qui rend plus précis et robustes les résultats de la détection et du suivi de mouvement sur des images vidéo [MC02]. Une stratégie fondée sur la logique floue [21] [AvKLM01] permet de fusionner les informations fournies par tous les descripteurs pour définir les positions dans l'espace 3D de tous les éléments modélisés de la salle.

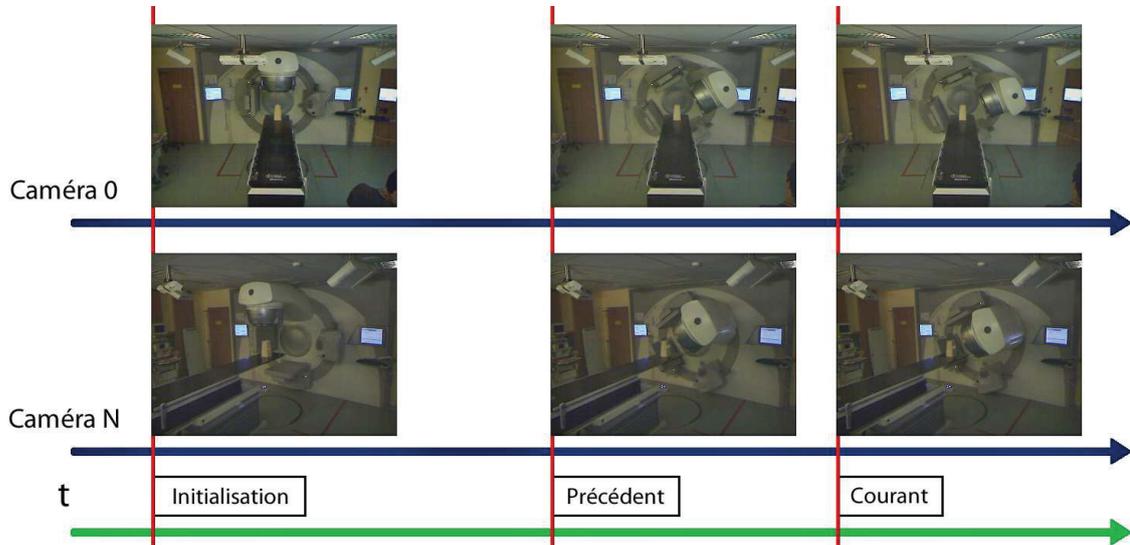


FIGURE 4.3: Principe temporel de la méthode de suivi de mouvement : on considère que les deux environnements sont bien corrélés à l'initialisation et, à tout instant, les données acquises à l'instant précédent sont utilisées pour calculer la pose à l'instant courant et ensuite mettre à jour le modèle.

4.3.1 Algorithme global de la méthode

Le schéma de la Fig. 4.4 illustre les étapes de l'algorithme global de la méthode de suivi de mouvement des éléments par le calcul de leur pose à tout instant. Comme

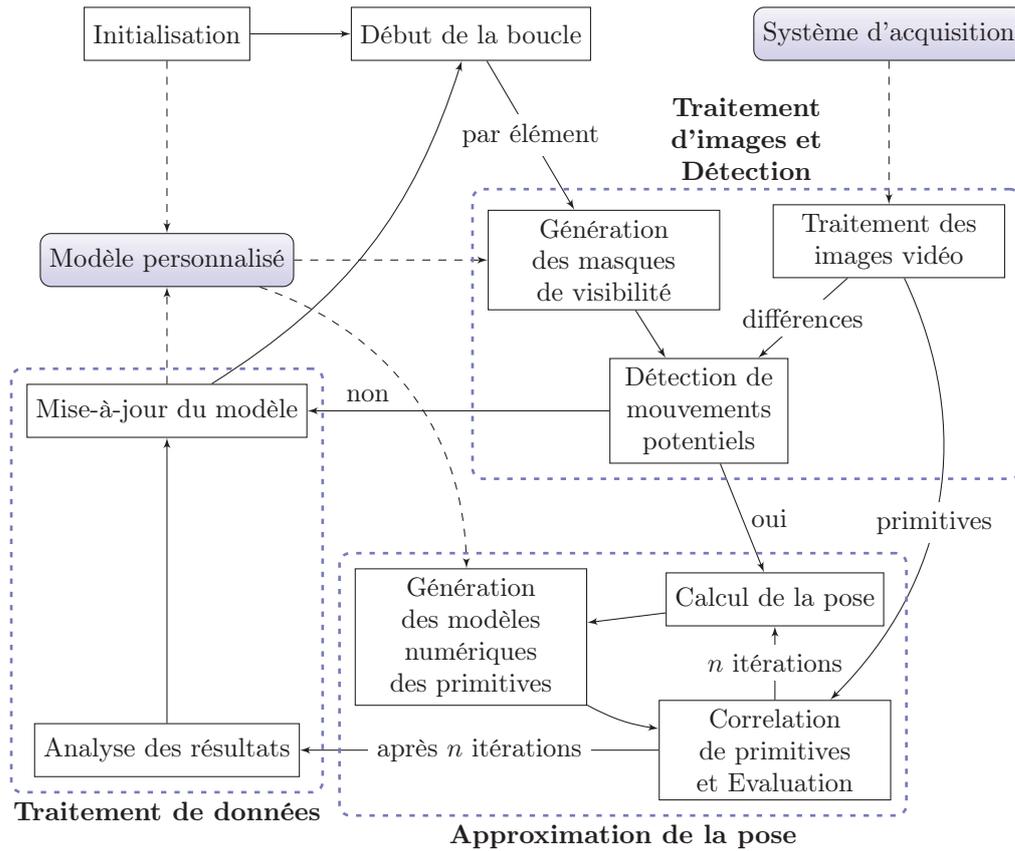


FIGURE 4.4: Méthode de calcul de la pose des éléments.

on peut l'observer, une phase d'initialisation est nécessaire afin de fournir un point de départ dans lequel les deux environnements sont bien corrélés. Cette phase a déjà été présentée à la section 3.3.6.

Après l'initialisation, on rentre dans une boucle globale de traitement dont chaque itération consiste en trois phases consécutives :

1. La phase de **traitement d'images et détection** se focalise sur la détection de mouvements car, afin de réduire le volume de calculs, la pose de l'élément est calculée uniquement si l'élément s'est effectivement déplacé.
2. La phase d'**approximation de la pose** réalise la recherche de la pose optimale, dans l'espace de solutions, qui corrèle au mieux les images réelles et virtuelles.
3. La phase de **traitement de données** est chargée de l'intégration, dans le modèle numérique, des informations acquises dans les phases précédentes.

Les itérations de cette boucle globale sont synchronisées avec la fréquence d'acquisition des caméras, et donc le temps t . Pour chaque acquisition, on calcule donc la pose de tous les éléments de manière séquentielle. La phase d'approximation de la pose est à son tour composée d'une boucle interne. À chaque passage dans cette phase, la boucle interne est parcourue n fois, une fois pour chaque élément d'intérêt contenu dans le modèle de la scène (certains éléments modélisés ne seront pas suivis, comme les objets statiques : murs, mobilier, etc.).

4.3.2 Modèle temporel du flux vidéo

L'extraction des éléments dans les images d'une séquence vidéo d'une scène filmée est la première étape de nombreuses applications dans le domaine de la vision par ordinateur. Une méthode très courante est la segmentation par soustraction de l'image courante avec une image de référence. La seuillage de cette différence permet d'obtenir un image binaire dont les pixels de valeur maximum correspondent aux éléments du premier plan.

L'image de référence utilisée représente généralement une image de la scène statique où aucun des éléments d'intérêt n'est présent (image du fond ou arrière plan). Cependant, l'obtention d'une image de ce type, dans une salle de traitement en radiothérapie, est très difficile à obtenir car il y a des éléments toujours présents (accélérateur, table de traitement, etc.). Une méthode simple, comme le moyennage temporel des valeurs des pixels, ne suffit pas à obtenir une image de référence. De plus, une telle image de référence ne peut pas être constante du fait de la variation des conditions et des paramètres de la salle de traitement.

Nous avons donc décidé de représenter cette image de référence par un modèle temporel des valeurs des pixels des images dans le flux vidéo. Cette modélisation est basée sur les propriétés statistiques des pixels afin de définir les valeurs possibles de chacun (Fig. 4.5). Plus précisément, ce qu'on cherche à modéliser c'est l'histogramme temporel de chaque pixel, qui correspond au nombre d'occurrences des différentes valeurs d'un pixel dans le temps.

La modélisation par mélange de gaussiennes a été présentée comme une solution robuste aux bruits, permettant de gérer les problèmes liés aux variations de lumière, de mouvement continu dans l'arrière plan (par exemple, feuilles d'arbres bougeant avec le vent), le déplacement de plusieurs objets, et certains changements arbitraires [SG99, Ziv04, Lee05]. Dans ce modèle, chaque gaussienne représente un état du pixel qui correspond généralement à l'un des éléments de la scène, sauf l'une des gaussiennes qui correspond à l'arrière plan (Fig. 4.6). Lee et al. [Lee05] proposent de définir l'arrière plan, non pas par une seule gaussienne, mais par l'ensemble des gaussiennes les plus présentes. Ainsi, un élément qui a arrêté de se déplacer sera intégré à l'arrière plan. Nous ne parlerons pas de l'aspect théorique des outils mathématiques utilisés pour la construction du modèle, qui sont bien détaillés dans [SG99].

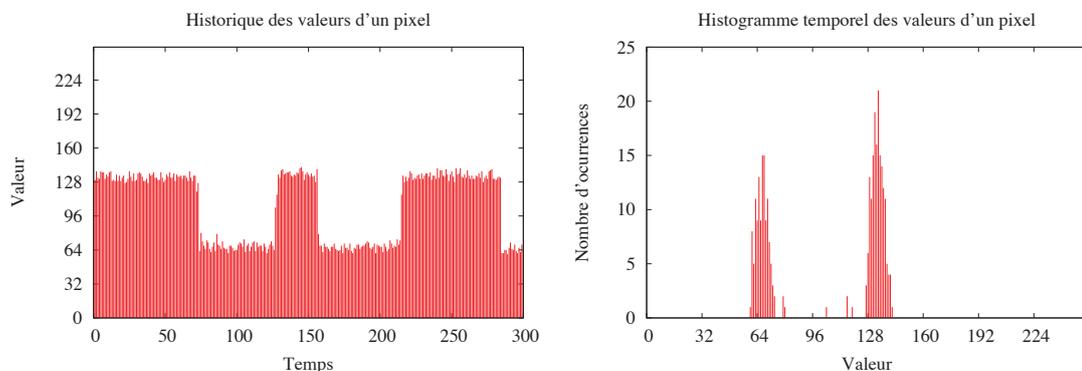


FIGURE 4.5: Exemple des propriétés statistiques d'un pixel.

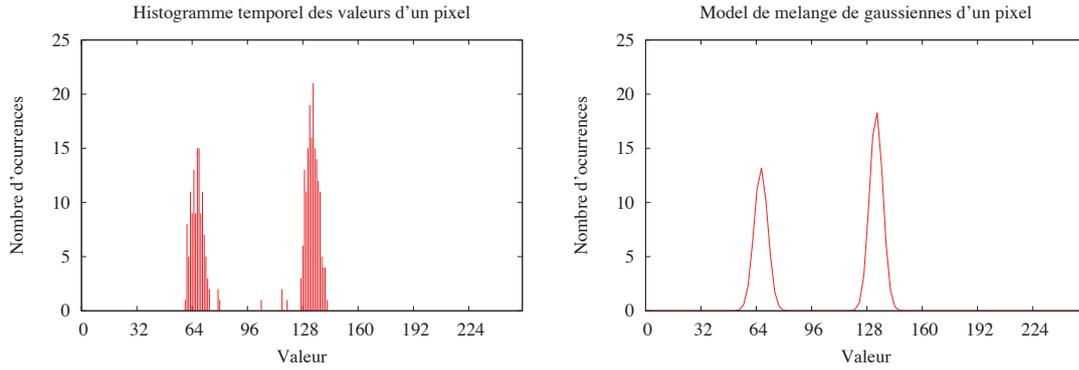


FIGURE 4.6: Exemple de modèle de mélange de gaussiennes d'un pixel.

L'implémentation du modèle que nous avons adopté est basée sur la proposition de Lee et al. [Lee05]. Une gaussienne, d'indice k dans le modèle, est définie par trois paramètres variant à chaque instant t :

- le poids $w_{(k,t)}$ de la gaussienne, qui représente la proportion des données associées (plus les valeurs sont présentes, plus le poids de la gaussienne est important),
- la valeur moyenne $\mu_{(k,t)}$ sur laquelle la gaussienne est centrée,
- la variance $\sigma_{(k,t)}^2$ des valeurs de la gaussienne.

Le modèle par mélange de gaussiennes sera défini par un ensemble de gaussiennes. Au départ, cet ensemble est vide. À chaque instant t , on cherche dans le modèle la gaussienne à laquelle appartient la nouvelle valeur acquise, et on met à jour les paramètres de toutes les gaussiennes. Si la valeur n'appartient à aucune des gaussiennes, on ajoute une nouvelle gaussienne au modèle. Le critère d'appartenance est donné par :

$$|\mu_{(k,t-1)} - PixelValue_t| < \lambda \times \sigma_{(k,t-1)} \quad (4.6)$$

où $PixelValue_t$ est la nouvelle valeur et λ est une valeur fixe prédéfinie (choisie en fonction des plages de normalité, ou intervalles de confiance). Pour des questions de simplicité de calcul, une nouvelle gaussienne aura les valeurs suivantes :

$$\mu_{(k,t)} = PixelValue_t, \quad \sigma_{(k,t)} = 10 \text{ et } w_{(k,t)} = 0.01$$

les deux dernières valeurs étant définies d'après les résultats des tests réalisés. Pour le critère d'appartenance on a considéré la valeur $\lambda = 3$ car 99.7% de la population est dans l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$. Le paramètre $w_{(k,t)}$ de toutes les gaussiennes est mis à jour suivant l'équation suivante :

$$w_{(k,t)} = (1 - \alpha)w_{(k,t-1)} + \alpha M_{(k,t)} \quad (4.7)$$

où $\alpha = 0.001$ est une constante d'apprentissage (valeur étant également définie d'après les résultats des tests réalisés), et $M_{(k,t)}$ prend le valeur suivante : 1 si la nouvelle valeur acquise appartient à la gaussienne d'indice k , et 0 sinon.

Ensuite, uniquement pour la gaussienne à laquelle appartient la nouvelle valeur (sauf si cette gaussienne vient d'être insérée dans le modèle), les paramètres $\mu_{(k,t)}$ et

$\sigma_{(k,t)}^2$ sont mis à jour suivant les équations suivantes :

$$\mu_{(k,t)} = (1 - \alpha)\mu_{(k,t-1)} + \alpha PixelValue_t \quad (4.8)$$

$$\sigma_{(k,t)}^2 = (1 - \alpha)\sigma_{(k,t-1)}^2 + \alpha(\mu_{(k,t)} - PixelValue_t)(\mu_{(k,t)} - PixelValue_t) \quad (4.9)$$

À chaque instant t , après la mise à jour des paramètres, les gaussiennes du modèle sont triées et renumérotées (indice k) en fonction du rapport $\mu_{(k,t)}/\sigma_{(k,t)}$. Ensuite, l'arrière plan est défini par les G premières gaussiennes telles que :

$$G_t = \underset{g}{arg \min} \left(\sum_k^g w_{(k,t)} > F \right) \quad (4.10)$$

où F est une valeur fixe prédéfinie (choisie en fonction de la proportion de données minimales que l'on veut associer au fond).

Le modèle permet d'obtenir ensuite une image binaire $imgForeground_{(i,t)}$, dont les pixels ont une valeur égal à : une valeur fixe différente de zéro (255 par exemple) si la nouvelle valeur du pixel appartient à une gaussienne k tel que $k > B_t$ (gaussienne n'appartenant pas l'arrière plan) ; 0 sinon.

Dans le but de détecter un mouvement, on souhaite déterminer l'identification des pixels dont la valeur a varié de manière significative. Toujours dans l'idée que chaque gaussienne est associée à un élément, ou à l'arrière plan, on cherche les pixels qui ont changé d'état. Pour ceci, on associe à chaque pixel un paramètre v supplémentaire qui enregistre la valeur de l'indice k de la dernière gaussienne visitée, c'est-à-dire, la gaussienne à laquelle appartient la valeur $PixelValue_{t-1}$. À l'aide de ce paramètre on peut obtenir une image binaire $imgDiff_{(i,t)}$, dont les pixels ont une valeur égal à : une valeur fixe différente de zéro (255 par exemple) si $v \neq k$, où k est l'indice de la gaussienne à laquelle appartient la valeur $PixelValue_t$ (la valeur de l'indice est considéré avant que les gaussiennes soient triées et renumérotées) ; 0 sinon.

4.3.3 Détection de mouvements

La première phase de la méthode s'intéresse à identifier les éléments qui se sont déplacés. Pour ceci (en reprenant la nomenclature introduite à la section 3.3.4), la méthode génère une image $VSimg_{(i,t,j,k_v)}$ ⁵ correspondant au descripteur « plus petit volume englobant » $Desc_{(j,k_v)}$. Cette image représente notamment la silhouette de l'élément à l'emplacement où l'élément était présent au temps $t - 1$ (au niveau image). Les pixels de cette image ont une valeur égale à la valeur de profondeur (z-buffer) si le pixel appartient à la silhouette, 0 sinon.

Si l'on considère l'ensemble de toutes les images $VSimg_{(i,t,j,k_v)}$ pour tous les éléments, il est possible d'obtenir une image binaire dont tous les pixels de valeur non nulle appartiennent à la partie visible de l'élément dans l'image vidéo. Cette image binaire, que nous avons appelée « masque de visibilité » et notée $imgVis_{(i,t,j)}$, est créée de la manière suivante : étant donné un élément $Elem_{j_c}$, pour chaque

5. pour rappel : i correspond à la caméra Cam_i , t à l'instant courant, j à l'élément $Elem_j$ et k est l'identifiant du descripteur au sein de l'élément.

pixel p de coordonnées (u, v) dans l'image $VSimg_{(i,t,j_c,k_v)}$ ayant une valeur différente de zéro, on vérifie dans chacune des autres images $VSimg_{(i,t,j,k_v)}$ que le pixel de coordonnées (u, v) a une valeur plus grande que la valeur de p . Si c'est le cas, le pixel de coordonnées (u, v) dans l'image $imgVis_{(i,t,j)}$ prend une valeur fixe différente de zéro (255 par exemple), ou la valeur 0 sinon. Pour tous les pixels p de coordonnées (u, v) de l'image $VSimg_{(i,t,j_c,k_v)}$ ayant une valeur égale à zéro, le pixel de coordonnées (u, v) dans l'image $imgVis_{(i,t,j)}$ prend la valeur 0. Ceci peut être représenté par l'équation suivante :

$$imgVis_{(i,t,j_c)(u,v)} = \begin{cases} Val_{max} & \text{si } Val_{j_c} \neq 0 \text{ et } Val_{j_c} < \min_{\forall j/j \neq j_c, Val_j \neq 0} Val_j \\ 0 & \text{sinon} \end{cases} \quad (4.11)$$

où Val_{max} est une valeur fixe prédéfinie différente de 0, et $Val_j = VSimg_{(i,t,j,k_v)(u,v)}$.

Comme les zones de plus haut gradient dans les images sont plus représentatives pour la détection temporelle des mouvements, nous nous intéressons aux contours visibles de l'élément. On se focalise donc sur le calcul de la proportion de pixels correspondant aux contours visibles de l'élément, calculée suivant l'équation :

$$ValDetect_{(i,t,j_c)} = \frac{|(imgVis_{(i,t,j_c)} \cap Rprim_{(i,t,j_c,k_e)}) \cap D|}{|imgVis_{(i,t,j_c)} \cap Rprim_{(i,t,j_c,k_e)}|} \quad (4.12)$$

où k_e est l'indice du descripteur « contours » de $Elem_{j_c}$ (autrement dit, $Rprim_{(i,t,j_c,k_e)}$ est l'ensemble des pixels appartenant aux contours extraits dans l'image vidéo courante) et D est l'ensemble de pixels non nul de $imgDiff_{(i,t)}$ (image obtenue grâce au modèle temporel du flux vidéo).

La valeur obtenue est ensuite comparée à un seuil fixe prédéfini afin d'établir si l'élément s'est déplacé ou non, et la pose d'un élément sera calculée uniquement si on a détecté un déplacement.

4.4 Fonction de dissimilarité

La deuxième phase de l'algorithme, « approximation de la pose » (Fig. 4.4), est composée d'une deuxième boucle pour le calcul de la pose d'un élément. Dans cette boucle, les images réelles acquises et les images virtuelles générées par ordinateur sont comparées dans le but de corréliser les environnements réel et virtuel. Les informations contenues dans ces images sont confrontées en fonction des descripteurs inclus dans le modèle de chaque élément, en comparant les primitives réelles et virtuelles. Les primitives réelles résultent de la segmentation des images vidéo, en utilisant des outils existants pour le type de primitive qu'on souhaite traiter. Par exemple, pour les contours, nous utilisons un filtre de type Canny, pour les descripteurs colorimétriques, nous exploitons l'apprentissage d'histogramme HSV (Teinte Saturation Valeur) pour obtenir une carte de probabilités, etc.. Les primitives virtuelles sont obtenues par la simulation d'une acquisition à partir des modèles numériques des caméras et du descripteur associé à la primitive. Ces primitives sont représentées par les ensembles $Rprim_{(i,t,j,k)}$ et $Vprim_{(i,t,j,k)}$ (cf. section 3.3.4).

L'implémentation de la mesure du critère de similarité pour le recalage 3D est souvent définie dans la littérature par une fonction de dissimilarité. Une fonction beaucoup utilisée est la distance de chanfrein [Bro88], qui est basée sur les contours. Si l'on considère deux contours c_a et c_b représentant des suites des pixels, la distance de c_a à c_b est donnée par :

$$DistanceChanfrein(A, B) = \frac{1}{|A|} \sum_{a_i \in A} \min_{b_j \in B} d(a_i, b_j) \quad (4.13)$$

où A est l'ensemble de pixels contenus dans c_a (de même pour B et c_b) et $d(p_1, p_2)$ est la distance euclidienne entre pixels p_1 et p_2 .

Une autre fonction très courante est la distance de Hausdorff [HKKR93], basée également sur les contours, mais qui calcule la valeur maximale des distances euclidiennes entre chaque pixel p_1 du contour c_a et le pixel p_2 du contour c_b le plus proche de p_1 . Cette distance est donné par :

$$DistanceHausdorff(A, B) = \max_{a_i \in A} (\min_{b_j \in B} d(a_i, b_j)) \quad (4.14)$$

où A , B et $d(p_1, p_2)$ ont la même définition que dans l'Eq. 4.13.

Le principal inconvénient des fonctions basées sur les contours est leur sensibilité aux données aberrantes⁶, ce qui arrive souvent dans notre cas du fait de la similarité en terme de couleurs entre les objets présents dans la salle de traitement (machines, mobilier, et autres). Pour résoudre à ce problème, il est possible de réaliser un pré-traitement sur les données, afin de ne garder que les plus représentatives. Une autre approche consiste à prendre en considération l'espace occupé par l'objet dans l'image. Ainsi, la notion de non recouvrement entre surfaces [OH99b] permet d'évaluer la différence entre deux silhouettes. Horain and Bomb [HB02] ont proposé d'évaluer le rapport entre le nombre de pixels dans la région d'intersection et le nombre de pixel dans la région de l'union des silhouettes, donné par :

$$DistanceNonRecouvrement(S_a, S_b) = \frac{|S_a \cup S_b| - |S_a \cap S_b|}{|S_a \cup S_b|} \quad (4.15)$$

où S_a et S_b sont deux ensemble de pixels, chacun contenant les pixels d'une des silhouettes considérées.

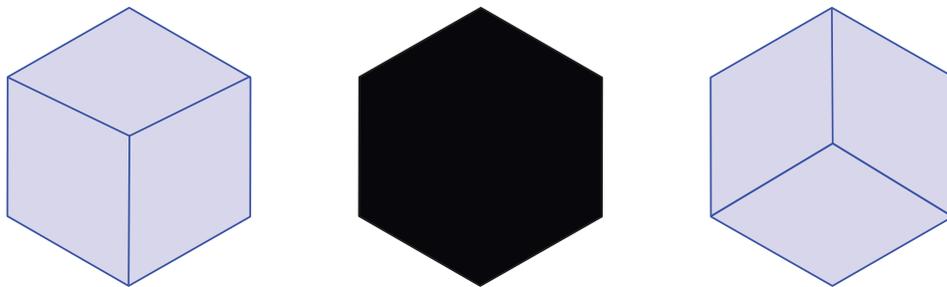


FIGURE 4.7: Exemple illustrant le problème des silhouette et l'avantage d'exploiter les contours internes (gauche : vue d'en haut, centre : silhouette, droit : vue d'en bas).

⁶. une donnée aberrante est une donnée sensiblement distante des autres données de l'ensemble dans laquelle elle se produit.

Cependant, la notion de non-recouvrement présente tous les inconvénients liés aux silhouettes, notamment le fait qu’une silhouette peut être le résultat de la projection de l’élément avec des poses différentes, incluant les auto-occultations (Fig. 4.7). Dans le but d’obtenir plus de robustesse, Ben Henia et al. [BHB10] ont proposé une fonction de dissimilarité qui combine la notion de non-recouvrement et la distance de chanfrein. Cette fonction considère la région représentant la différence entre deux régions issues de l’union et de l’intersection des silhouettes, et elle évalue la distance des pixels dans cette région par rapport aux contours de l’une des silhouettes. La distance de S_a à S_b est donnée par :

$$DistanceMixte(S_a, S_b) = \sum_{p \in SNR} dm(p, Contour_{S_b}) \quad (4.16)$$

où :

- S_a et S_b ont la même définition que dans l’Eq. 4.15,
- SNR est l’ensemble résultat de $((S_a \cup S_b) - (S_a \cap S_b))$ (Surface de Non Recouvrement),
- $Contour_{S_b}$ est l’ensemble des pixels du contour de la silhouette dont les pixels sont contenus dans l’ensemble S_b ,
- $dm(p, C)$ est la distance euclidienne entre le pixel p et celui le plus proche dans $Contour_{S_b}$.

Bien que cette fonction donne de meilleurs résultats pour le suivi des mouvements de la main, par rapport aux fonctions présentées précédemment, elle n’offre pas de solution à la problématique des silhouettes.

Puisqu’aucune fonction de la littérature ne donnent pas à elle seule d’assez bons résultats dans le cadre de nos travaux, nous avons dû de recourir à une solution hybride, sous la forme d’une combinaison de deux fonctions : distance de chanfrein et notion de non-recouvrement. Cette fonction exploite les contours de l’élément, au lieu d’uniquement les contours de la silhouette, afin de mieux prendre en compte la topologie de l’élément. Notre fonction de dissimilarité définit la distance entre deux élément $Elem_a$ et $Elem_b$ par :

$$\rho(Elem_a, Elem_b) = \left(1 - \frac{|S_a \cap S_b|}{|S_a \cup S_b|}\right) \times \frac{1}{|C_a|} \times \sum_{p_a \in C_a} \min_{p_b \in C_b} d(p_a, p_b) \quad (4.17)$$

où, S_j est ensemble de pixels contenus dans la silhouette de $Elem_j$ ($j = \{a, b\}$), C_j est l’ensemble des pixels contenus dans les contours de $Elem_j$, et $d(p_a, p_b)$ est la distance euclidienne entre les coordonnées des pixels p_a et p_b . Comme on peut le remarquer (en regardant également Eq. 4.13 et Eq. 4.15), cette fonction comporte deux parties qui évaluent la distance entre les éléments selon deux cas mutuellement exclusifs.

Dans le cadre du recalage, l’utilisation des fonctions présentées permet d’évaluer la distance entre chacune des poses dans l’ensemble de solutions par rapport à une pose « cible », et ensuite de les comparer pour garder la meilleure. La pose cible $Pose3D_T$ est la pose de l’élément que l’on cherche à calculer (ou à approcher), et les informations associées à cette pose sont extraites des images réelles. Dans notre cas, l’élément $Elem_a$ représente l’élément réel que l’on suit dans les images et

l'élément $Elem_b$ représente son modèle numérique. Le modèle numérique est déplacé dans l'espace de la scène en fonction des poses contenues dans l'espace des solutions de l'élément. La silhouette et les contours du modèle numérique sont obtenus par projection sur le plan image, alors que la silhouette et les contours de l'élément réel sont obtenus par extraction sur les images vidéo.

Nous allons considérer deux poses à comparer $Pose3D_A$ et $Pose3D_B$ contenues dans l'espace de solutions et différentes de la pose cible $Pose3D_T$, afin de mieux comprendre le comportement de notre fonction :

1. lorsque la silhouette du modèle numérique, placée dans la $Pose3D_A$ ou dans la $Pose3D_B$, a une intersection vide avec la silhouette cible (voir Fig. 4.8(a)), la première partie de la fonction Eq. 4.17, correspondant à la notion de non-recouvrement

$$\left(1 - \frac{|S_a \cap S_b|}{|S_a \cup S_b|}\right),$$

prend la valeur 1, car $|S_a \cap S_b| = 0$, et la fonction se comporte comme une fonction de chanfrein :

$$\rho(Elem_a, Elem_b) = \frac{1}{|C_a|} \times \sum_{p_a \in C_a} \min_{p_b \in C_b} d(p_a, p_b)$$

ce qui donne plus de poids à la translation du modèle numérique vers la pose cible,

2. dans le cas contraire (Fig. 4.8(b)), la première partie de la formule apporte des informations plus précises sur la rotation, car elle porte sur le nombre de pixels et non sur une moyenne de distances.

L'utilisation de la silhouette de l'élément permet de surmonter les problèmes liés aux occultations partielles et aux variations d'éclairage. L'utilisation des contours de l'élément, au lieu d'uniquement les contours de la silhouette de l'élément, permet de garantir la robustesse lors d'occultations partielles et d'auto-occultations, et lorsque la silhouette est symétrique par rapport à un plusieurs degrés de liberté de l'élément.

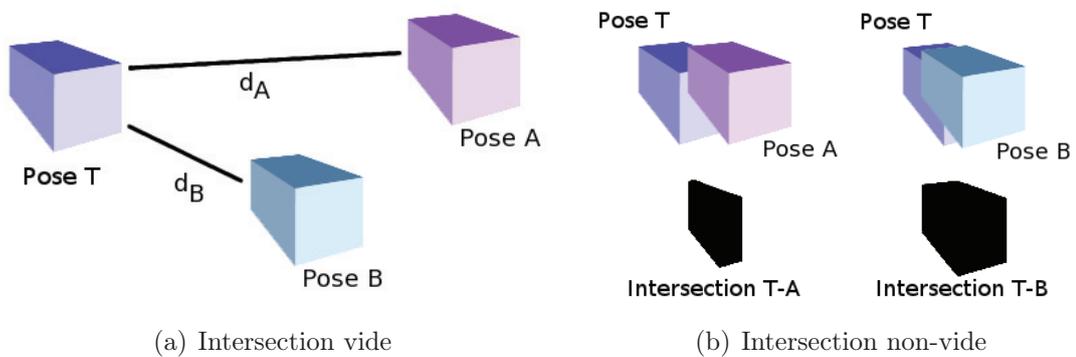


FIGURE 4.8: Illustration de la fonction de dissimilarité.

4.4.1 Carte de distances

Le première étape des méthodes de recalage utilisant des fonctions de dissimilarité qui incluent des mesures de distance euclidienne, est le calcul d'une carte de distances [Dan80, Bor86]. Cette carte prend la forme d'une image avec les mêmes dimensions que les images exploitées (images vidéo dans notre cas), et elle est créée à partir des informations extraites de ces images. Chaque pixel p de cette image a comme valeur la distance entre lui et le pixel le plus proche contenu dans l'ensemble de pixels associé aux informations extraites (généralement des contours). Cette valeur correspond à l'équation suivante :

$$CarteDist(C)_{(u,v)} = \min_{c \in C} da(p_{(u,v)}, c) \quad (4.18)$$

où C est l'ensemble de pixels du contour de l'élément réel, p est un pixel dans la carte de distances et u et v sont ses coordonnées, et $da(p, c)$ est une approximation de la distance euclidienne entre les coordonnées des pixels p et c .

L'utilisation d'une carte de distances est justifiée par l'existence d'algorithmes de calcul performants [MQR03, FH04], avec une complexité en $O(n)$, par rapport au calcul classique de la distance à partir de chaque pixel qui a une complexité en $O(n^2)$. Il est également possible de normaliser cette image afin d'obtenir une image avec des valeurs entières comprises entre 0 et 255, pour effectuer les opérations uniquement avec des entiers. Certains algorithmes proposent également de calculer la distance euclidienne, sur une autre métrique que les coordonnées pixels, permettant d'obtenir des valeurs entières approchées. De plus, cette carte est générée une seule fois au début du processus de recherche dans l'espace de solutions à un instant t , car les informations réelles restent inchangées.

En utilisant Eq. 4.18, l'équation Eq. 4.17 devient :

$$\rho(Elem_a, Elem_b) = \left(1 - \frac{|S_a \cap S_b|}{|S_a \cup S_b|}\right) \times \frac{1}{|C_a|} \times \sum_{p_{(u,v)} \in C_a} CarteDist(C_b)_{(u,v)} \quad (4.19)$$

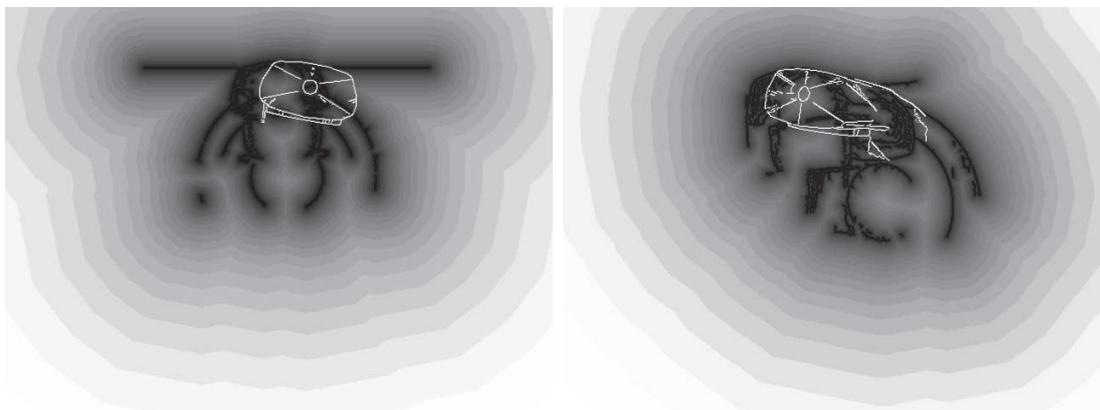


FIGURE 4.9: Exemple de cartes de distances d'un même élément sur des images de deux caméras. Les contours du modèle virtuel sont superposés afin d'illustrer les informations qui sont recherchées dans la carte.

4.5 Méthodologie générale du calcul de pose

Comme introduit précédemment, la fusion des données *a priori* issues des descripteurs de l'élément et des éléments dans son entourage permet d'améliorer la précision et la robustesse de la méthode de calcul de la pose. Nous avons proposé trois stratégies afin de permettre l'implémentation de ce principe de fusion :

1. l'utilisation de la logique floue pour la fusion de données hétérogènes et la prise en compte de la visibilité des descripteurs et de la dynamique de l'élément,
2. la réduction de l'espace de recherche à l'aide des contraintes définies dans le modèle numérique,
3. l'utilisation de files de priorité pour le calcul des poses de l'ensemble d'éléments et la réutilisation de certaines informations.

Plus précisément, la fonction de dissimilarité est utilisée pour chaque descripteur, et ensuite toutes les mesures obtenues sont fusionnées afin d'obtenir une valeur unique pour l'élément. Pour chaque descripteur $Desc_{(j,k)}$, nous considérons :

1. La silhouette, qui est contenue dans les ensembles $Rprim_{(i,t,j,k_s)}$ et $Vprim_{(i,t,j,k_s)}$ (cf. section 3.3.4). Suivant la même procédure de calcul du masque de visibilité d'un élément, présentée à la section 4.3.3, nous pouvons obtenir un masque de visibilité pour un descripteur. $Vprim_{(i,t,j,k_s)}$ est l'ensemble des pixels non nuls de ce masque. $Rprim_{(i,t,j,k_s)}$ est obtenu par extraction sur les images en utilisant les outils déjà définis et adaptés au type de descripteur et en exploitant des informations temporelles, notamment le masque de visibilité du descripteur à l'instant précédent $t - 1$.
2. Les contours, sont obtenus à partir de l'intersection du descripteur « contours » $Desc_{(j,k_e)}$ de l'élément et du masque de visibilité du descripteur k . De ce fait, le descripteur $Desc_{(j,k_e)}$ ne sera pas considéré dans la liste des descripteurs à évaluer. Les contours réels $CRprim_{(i,t,j,k)}$ et les contours virtuels $CVprim_{(i,t,j,k)}$ du descripteur $Desc_{(j,k)}$ sont donc obtenus par :

$$CRprim_{(i,t,j,k)} = Rprim_{(i,t,j,k)} \cap Rprim_{(i,t,j,k_e)} \quad (4.20)$$

$$CVprim_{(i,t,j,k)} = Vprim_{(i,t,j,k)} \cap Vprim_{(i,t,j,k_e)} \quad (4.21)$$

où k_e est l'indice associé au descripteur « contours » dans l'élément, et $k \neq k_e$.

En incorporant la notion de descripteur dans Eq. 4.19, on obtient le poids associé au $Desc_{(j,k)}$ dans la pose $Pose3D_{(j,t)_n}$ dans l'image de la caméra i à l'instant t :

$$PImgDesc_{(i,j,k,t)_l} = \left(1 - \frac{|V \cap R|}{|V \cup R|}\right) \times \frac{1}{|CV|} \times \sum_{p_{(u,v)} \in CV} CarteDist(CR)_{(u,v)} \quad (4.22)$$

où l est l'indice associé à la $Pose3D_{(j,t)_n}$ dans l'espace de solutions⁷, et

$$V = Vprim_{(i,t,j,k)}, R = Rprim_{(i,t,j,k)}, CV = CVprim_{(i,t,j,k)} \text{ et } CR = CRprim_{(i,t,j,k)}$$

7. en réalité, l'espace de solutions n'est pas un espace dénombrable. À chaque itération, on réalise un échantillonnage contenant les poses visitées lors du parcours de l'espace de solutions.

4.5.1 Évaluation des descripteurs pour la calcul d'une pose

Le principe de la logique floue repose sur l'utilisation d'ensembles flous pour permettre la fusion de données hétérogènes et la prise en compte de l'imprécision générée par l'acquisition et le traitement des données [Zad65]. Un ensemble flou est un ensemble dans lequel les éléments ont des degrés d'appartenance, contrairement aux ensembles théoriques classiques où l'appartenance a une valeur binaire. Un ensemble flou est défini par un ensemble classique d'éléments E et par une fonction d'appartenance $m : E \rightarrow [0, 1]$. L'appartenance de l'élément $x \in E$ dans l'ensemble flou (E, m) sera donnée par : non inclus, si $m(x) = 0$, totalement inclus, si $m(x) = 1$, ou inclus avec la probabilité (ou « taux de confiance ») d'appartenance donnée par $m(x)$.

Nous définissons deux ensembles flous correspondants à la visibilité (au niveau de l'image) et au déplacement dans l'espace 3D. Ces ensembles permettent d'associer une mesure de poids représentant un taux de confiance sur l'approximation calculée pour une pose cible.

Basé sur le principe que « plus un objet est visible dans sa totalité, plus on a des informations sur lui », le premier ensemble que nous avons défini mesure le rapport entre la partie visible de la silhouette du descripteur et la silhouette toute entière (Fig. 4.10). Cette mesure est donnée par :

$$VisDesc_{(i,j,k,t)} = \frac{|Vprim_{(i,t,j,k)}|}{|S_{(i,t,j,k)}|} \quad (4.23)$$

où $S_{(i,t,j,k)}$ est l'ensemble de pixels contenus dans la silhouette du descripteur (calculée en considérant que le descripteur est le seul élément dans la scène, donc entièrement visible), et $|Vprim_{(i,t,j,k)}|$ a une valeur entière entre 0 et $|S_{(i,t,j,k)}|$.

En regroupant tous les valeurs des différentes vues pour chaque descripteur, on obtient le poids associé à un descripteur dans une pose l à l'instant t par :

$$PoidDesc_{(j,k,t)_l} = \prod_i^m (2 - VisDesc_{(i,j,k,t)}) \times PImgDesc_{(i,j,k,t)_n} \quad (4.24)$$

où m est le nombre de caméras dans le système d'acquisition.

4.5.2 Pondération de la pose

Le deuxième ensemble flou qui a été défini porte sur la dynamique de l'élément, afin de prendre en compte ses caractéristiques mécaniques. Lorsqu'on s'intéresse au suivi de mouvement sur des flux vidéo, on peut rapidement s'apercevoir que le bruit contenu dans les images peut rendre « vraie » la détection du mouvement alors que l'élément ne s'est pas réellement déplacé. L'exemple le plus courant est le changement d'éclairage sur l'élément, qui peut être dû à la lumière ambiante ou à l'ombre générée par un autre élément, et qui peut donner comme résultat le calcul d'un très petit déplacement par des effets d'approximation. Cet ensemble flou mesure la faisabilité du déplacement calculé en fonction des paramètres mécaniques de l'élément (Fig. 4.11). Plus précisément, on cherche à évaluer si l'élément est

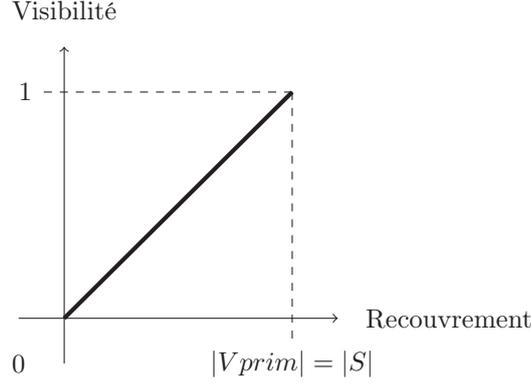


FIGURE 4.10: Ensemble flou pour la visibilité.

capable de réaliser un déplacement en-dessous d'un seuil prédéfini dans un pas de temps Δt . Autrement dit, l'élément doit réaliser un déplacement d'amplitude assez importante pour qu'on puisse conclure qu'il s'est réellement déplacé.

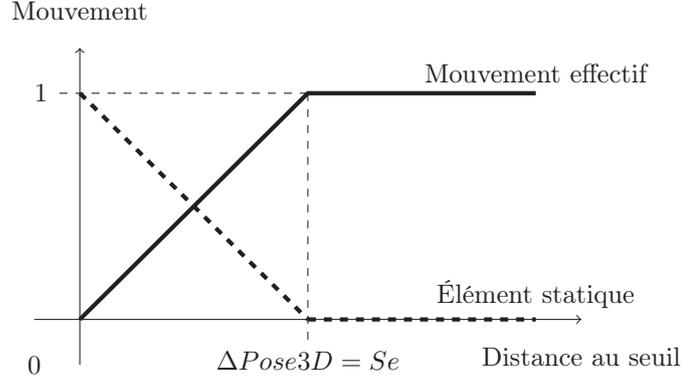


FIGURE 4.11: Ensembles flous pour le mouvement.

Le vecteur $\Delta Pose3D_{(j,t)}$ (Eq. 4.5) est comparé à un vecteur seuil fixe $Se_j = (T_{x_s}, T_{y_s}, T_{z_s}, R_{x_{0_s}}, R_{y_{1_s}}, R_{z_{2_s}})$ dont toutes les valeurs sont positives, et selon le mécanisme suivant :

- en dessous du seuil : si les valeurs absolues de tous ses composants sont strictement inférieures à leurs homologues respectifs dans Se_j ,
- au-dessus du seuil : sinon.

En effet, pour chaque paramètre on définit une valeur de seuil au-dessus de laquelle le déplacement est effectif. Il suffit que l'amplitude d'un seul paramètre soit assez proche de cette limite, ou plus grande, pour conclure à un déplacement. Le rapport $|Param_{(\Delta Pose3D_{(j,t),r})}| / Param_{(Se_j,r)}$ (où $Param_{(V,r)}$ est la composante d'indice r dans le vecteur V) est donc un discriminant adéquat. Nous utilisons ensuite l'opérateur OU en logique floue, défini à l'aide de la fonction max , pour fusionner les informations de tous les paramètres :

$$DiffVec_{(T_{r(j,t)}, Se_j)} = \max_r \left(\frac{|Param_{(\Delta Pose3D_{(j,t),r})}|}{Param_{(Se_j,r)}} \right) \quad (4.25)$$

Cette fonction permet de définir les fonctions d'appartenance de deux états qui

ont été définis dans cet ensemble : *statique* et *mouvement effectif*. Leur valeurs sont données par la mesure des différences entre $\Delta Pose3D_{(j,t)}$ et Se_j , et définies par :

$$Statique_{(j,t)} = \begin{cases} DiffVec_{(Tr_{(j,t)}, Se_j)} & \text{si } \Delta Pose3D_{(j,t)} \text{ est} \\ & \text{en dessous de } Se_j, \\ 1 & \text{sinon.} \end{cases} \quad (4.26)$$

$$Mouvement_{(j,t)} = 1 - Statique_{(j,t)} \quad (4.27)$$

La dynamique de l'élément est mesurée par la « certitude » que le mouvement calculé est un mouvement réalisable physiquement et géométriquement (Fig. 4.12). Les états *statique* et *mouvement effectif* sont donc fusionnés par l'opérateur *OU Exclusif* en logique floue, dont la formule est donnée par :

$$XOR(A, B) = A + B - 2 \times \min(A, B) \quad (4.28)$$

La mesure de la dynamique de l'élément est donc donnée par l'équation :

$$Dynamique_{(j,t)} = XOR(Statique_{(j,t)}, Mouvement_{(j,t)}) \quad (4.29)$$

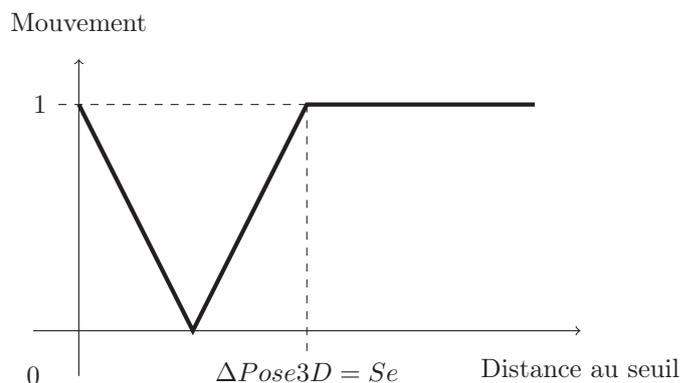


FIGURE 4.12: Ensemble flou pour la dynamique.

La pose d'un élément est évaluée à partir des descripteurs avec un taux de confiance en dessous d'un seuil prédéfini, afin de ne considérer que ceux qui sont les plus discriminants :

$$\varphi_{(j,k,t)_n} = \begin{cases} PoidDesc_{(j,k,t)_n} & \text{si } PoidDesc_{(j,k,t)_n} < SDesc \\ 1 & \text{sinon} \end{cases} \quad (4.30)$$

Enfin, le poids associé à une pose $Pose3D_{(j,t)}$ est donné par :

$$PoidPose3D_{(j,t)_n} = (2 - Dynamique_{(j,t)}) \times \prod_k^n \varphi_{(j,k,t)_n} \quad (4.31)$$

4.5.3 Algorithme d'optimisation

L'espace des solutions est exploré en utilisant la méthode d'optimisation non-linéaire proposée par Nelder et Mead [NM65]. Cette méthode est fondée sur les simplexes⁸ et elle réalise un parcours de l'espace des solutions dans des directions différentes autour d'un simplexe. La simplicité de la mise en œuvre, et le fait que la dérivée de la fonction à minimiser n'est pas nécessaire, justifient ce choix. Une autre raison est que cette approche pourrait être parallélisée afin de réduire les temps de calcul [LW07].

Dans l'algorithme 1 nous présentons l'implémentation utilisée, qui est basée sur l'implémentation présentée dans [PTVF92]. Il s'agit de traiter un ensemble de points, chacun contenant les paramètres de l'une des poses contenues dans l'espace des solutions. Ces points ont initialement des valeurs de poses proches de la pose de l'élément à l'instant $t - 1$. À chacun de ces points, on associe la valeur obtenue par l'évaluation de la fonction de dissimilarité en plaçant le modèle numérique de l'élément dans la pose définie par les paramètres du point. L'ensemble étant trié à chaque itération, il s'agit de remplacer le point ayant la valeur la plus grande par un nouveau point qui aura potentiellement une valeur plus basse.

Dans l'espace de recherche, le choix du nouveau point est fait dans la direction entre le point qu'on souhaite remplacer et un point représentant le centre géométrique de l'ensemble de points auquel on a déjà enlevé le point à remplacer. Dans l'exemple de la Fig. 4.13, cette direction est représentée par \vec{d}_k .

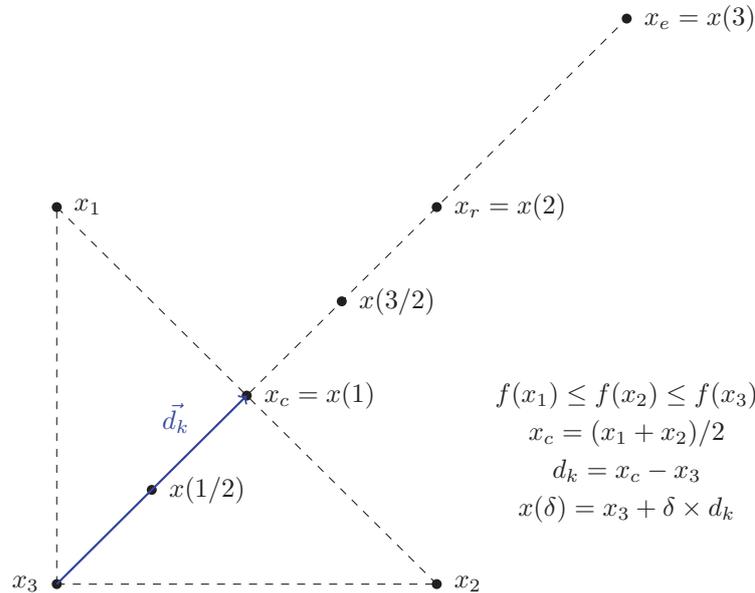


FIGURE 4.13: Exemple, en deux dimensions, des vecteurs pouvant être évalués dans une itération de la méthode de Nelder-Mead (même notation que dans l'algorithme 1).

Nous exploitons également les contraintes mécaniques des éléments, et en particulier les degrés de liberté comme suggéré par [OH99b], tels que définis dans le modèle numérique de la scène. En considérant toujours l'exemple de la Fig. 4.13, si le point à évaluer x_p ne respecte pas les contraintes définies (il est en dehors de

8. les simplexes sont des polytopes de $N + 1$ sommets dans un espace N -dimensionnel.

Algorithme 1 Méthode de Nelder-Mead Adaptée (*Downhill Simplex method*).

ENTRÉES:

- fonction $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ (fonction de dissimilarité dans notre cas)
- vecteurs du simplexe, affinement indépendents :
 $\{x_1^0, \dots, x_{n+1}^0 | x \in \mathbb{R}^n, \forall i \in \{1, \dots, n\}, f(x_i^0) \leq f(x_{i+1}^0)\}$
- précision requise pour les valeurs des vecteurs $\tau, \tau \in \mathbb{R}, \tau > 0$
- précision requise pour les valeurs de la fonction $\varepsilon, \varepsilon \in \mathbb{R}, \varepsilon > 0$
- nombre maximum d'itérations $\ell, \ell \in \mathbb{N}^+$

SORTIES: une approximation de la solution $x^* \in \mathbb{R}^n$

$k \leftarrow 0$

répéter

$x_c \leftarrow 1/n \sum_{i \leftarrow 1}^n x_i^k$

$d_k \leftarrow x_c - x_{n+1}^k$

$x_r \leftarrow x_{n+1}^k + 2d_k == 2x_c - x_{n+1}^k$

si x_r est en dehors l'espace de solution **alors**

$x_r \leftarrow CorrectionVecteur(x_r)$

fin si

si $f(x_r) < f(x_1^k)$ **alors**

$x_e \leftarrow x_{n+1}^k + 3d_k == 2x_r - x_c$

si x_e est en dehors l'espace de solution **alors**

$x_e \leftarrow CorrectionVecteur(x_e)$

fin si

si $f(x_e) < f(x_r)$ **alors**

$x_m \leftarrow x_e$

sinon

$x_m \leftarrow x_r$

fin si

sinon

si $f(x_1^k) \leq f(x_r) < f(x_n^k)$ **alors**

$x_m \leftarrow x_r$

sinon

si $f(x_n^k) \leq f(x_r)$ **alors**

si $f(x_{n+1}^k) \leq f(x_r)$ **alors**

$x_m \leftarrow x_{n+1}^k + 1/2d_k == 1/2(x_{n+1}^k + x_c)$

sinon

$x_m \leftarrow x_{n+1}^k + 3/2d_k == 1/2(x_r + x_c)$

si x_m est en dehors l'espace de solution **alors**

$x_m \leftarrow CorrectionVecteur(x_m)$

fin si

fin si

fin si

fin si

fin si

$x_{n+1}^{k+1} \leftarrow x_m, x_i^{k+1} \leftarrow x_i^k, i \in \{1, \dots, n\}$

$k \leftarrow k + 1$

Trier $\{x_i^k\}$ afin d'obtenir $f(x_i^k) \leq f(x_{i+1}^k), i \in \{1, \dots, n\}$

jusqu'à $\|d_k\| < \tau$ **ou** $f(x_0^k) < \varepsilon$ **ou** $\ell \leq k$

$x^* \leftarrow x_m$

l'espace de recherche), alors on l'échange avec un point x'_p , le plus proche de x_p , qui appartient à l'espace de solutions et qui respecte l'équation $x(\delta) = x_3 + \delta \times d_k$. C'est cette opération qui est réalisée par la fonction *CorrectionVecteur*(x) dans l'algorithme 1. Afin d'éviter les doublons, la dégénérescence du simplexe ou la perte de la propriété d'indépendance affine entre les points, une contrainte supplémentaire est posée lors du calcul du nouveau point x'_p . Cette contrainte correspond à une distance minimale fixe entre le point x'_p et tous les points du simplexe, ainsi que le point x_c représentant le centre géométrique. Enfin, le critère d'arrêt porte sur la précision souhaitée et le nombre d'itérations maximal accepté.

Avec tous les outils présentés, l'équation Eq. 4.4 est formalisée par :

$$\begin{aligned}
 & Pose3D_{(j,t)} = Pose3D_{(j,t)_o} , \text{ telle que} \\
 & PoidPose3D_{(j,t)_o} = \min_{Pose3D_{(j,t)_n} \in ETr} PoidPose3D_{(j,t)_n}
 \end{aligned}
 \tag{4.32}$$

où $ETr = \{ Pose3D_{(j,t)_n} / Pose3D_{(j,t)_n} \text{ a été visitée par l'algorithme du simplexe} \}$.

4.5.4 Utilisation des files de priorité

Nous avons vu précédemment que notre méthode de suivi de mouvement calcule la pose d'un élément en fonction de ses descripteurs. On s'intéresse maintenant à la manière dont l'ensemble des éléments est traité à chaque instant, suite à une nouvelle acquisition d'images vidéo. Bien que le calcul de la pose bénéficie de la notion de descripteur, dont l'imprécision est déjà prise en compte par les ensembles flous que nous avons définis, il se peut que dans certains cas les informations extraites des images vidéo ne soient pas suffisantes pour le calcul de la pose courante de l'élément. Heureusement, il est possible d'obtenir plus d'informations à partir du modèle numérique de la scène, notamment par les éléments dont connaît déjà la pose courante, soit parce qu'on l'a déjà calculée, soit parce qu'il ne s'est pas déplacé. De plus, même quand l'information est suffisante, les informations d'autres éléments peuvent accélérer le calcul de la pose de l'élément courant.

Pour ceci, nous avons mis en place une stratégie basée sur des files de priorité contenant les éléments dont on cherche à calculer la pose. Cette stratégie prend en compte le poids associé à une pose comme un critère permettant de décider s'il est nécessaire d'obtenir plus d'informations ou si l'élément courant peut fournir des informations aux autres éléments.

Les éléments contenus dans la file de priorité sont extraits par ordre décroissant en fonction de :

- leur importance : dans un sens sémantique, certains éléments doivent être traités en priorité (le meilleur exemple est le patient),
- leur dynamique : à partir des informations *a priori* dont on dispose sur l'élément on est capable de prévoir quels éléments seront les plus susceptibles de réaliser un déplacement (par exemple, il est plus probable qu'un individu réalise un déplacement inattendu par rapport à l'une des machines, et de même, les équipements robotisés auront plus tendance à se déplacer par rapport à ceux qui se sont pas robotisés),

- leur espace de liberté : prise en compte des autres éléments dans leur entourage, et avec qui l'élément courant pourrait interagir (par exemple, si un accessoire est ajouté au bras d'irradiation, les déplacements du bras d'irradiation feront que l'accessoire se déplacera également par rapport au référentiel global de la salle).

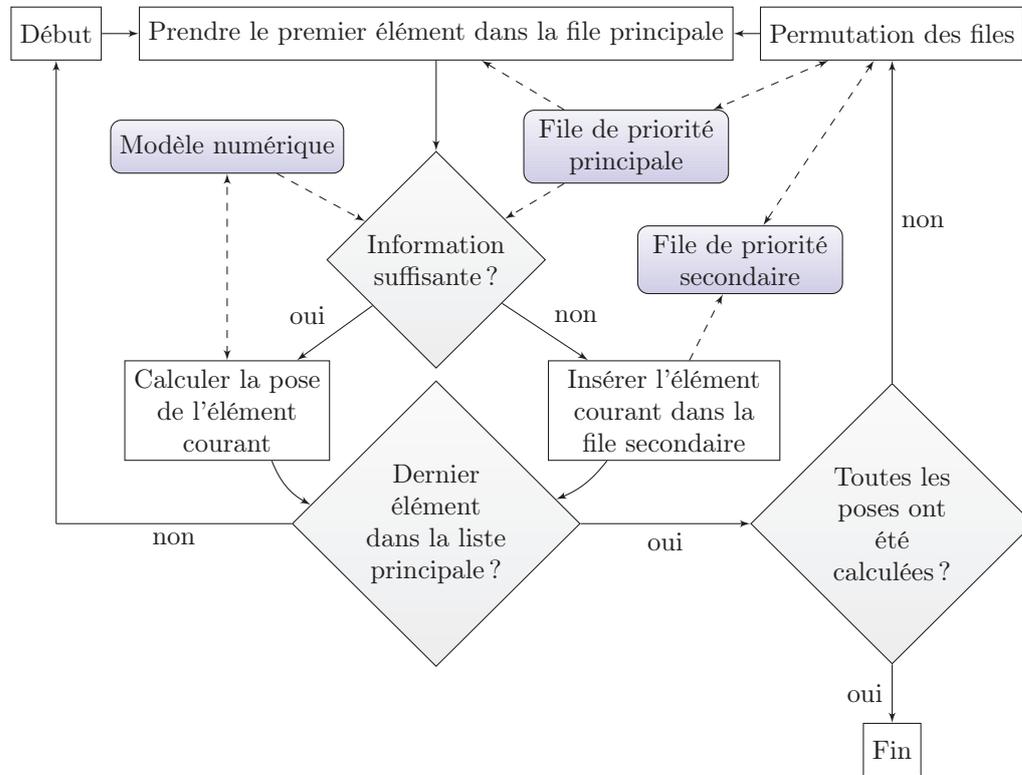


FIGURE 4.14: Méthode de calcul de la pose des éléments.

Le processus de traitement des files de priorité est organisé comme on peut le voir sur le diagramme de la Fig. 4.14. L'idée principale est de calculer la pose de chaque élément selon sa priorité, comme indiqué précédemment. La prise en compte du poids des poses déjà évaluées permet de pondérer leur contributions pour le calcul des poses des élément encore non réalisées. De la même manière, si l'information disponible n'est pas suffisante pour calculer la pose de l'élément courant (dernier extrait de la file), il est mis de coté et traité ultérieurement.

Le processus fonctionne avec deux files de priorité : une file principale et une file secondaire. À l'initialisation, la file principale contient tous les éléments pour lesquels on a détecté un déplacement. Lorsqu'on extrait un élément de cette file, le processus réalise une évaluation sur la disponibilité des informations pour le calcul de la pose. Si par l'évaluation on considère que les informations sont suffisantes, on engage le calcul de la pose de l'élément. Le cas échéant, on place l'élément dans une file de priorité secondaire. Cette évaluation est faite dans le nœud « Information Suffisante? » sur le diagramme de la Fig. 4.14, et elle est exécutée par rapport à deux critères qui sont :

1. La valeur du poids $PoidPose3D_{(j,t)}$ de la pose $Pose3D_{(j,t-1)}$ par rapport à la valeur d'un seuil associé à la file de priorité. En effet, il s'agit de réévaluer la

pose à l'instant précédent par rapport aux nouvelles images acquises, ce poids étant potentiellement différent à $PoidPose3D_{(j,t-1)}$. Le seuil associé à toute la file est incrémenté à chaque itération par un pas global, fixe et prédéfini, après avoir réalisé une permutation entre les deux files de priorité. Plus précisément, le poids d'une pose à une valeur entre 0 et une valeur maximale fixe $PoidMax$. Le pas avec lequel le seuil est incrémenté correspond à une fraction de cette valeur, et il est égal à $SeP = PoidMax/ItMax$, où $ItMax$ est le nombre maximum d'itérations du processus de traitement des files de priorité (valeur fixe et prédéfinie). Plus la valeur du seuil est grande, moins le critère est restreint : la valeur maximale $PoidMax$ ne correspond à aucune contrainte pour tous les éléments contenus dans la file de priorité. Lorsque le seuil global devient la valeur maximale, tous les éléments sont traités même si la valeur du poids de la pose suggère que la précision des résultats sera très incertaine.

2. La validité des opérateurs dans le modèle numérique, implémentant des notions sémantiques ou logiques, ou bien des interactions physiques ou géométriques (plus d'informations à la section 3.3.5). Par exemple, si on détecte un mouvement de la table de traitement quand le patient est déjà dessus, et qu'on calcule sa pose avant la pose du patient, alors l'opérateur liant la table et le patient peut être utilisé pour aider à calculer la pose du patient plus tard dans le processus (mais peut ne pas être suffisant si le patient à lui même bougé, par exemple).

Quand tous les éléments dans la file de priorité principale ont été traités, on permute les deux files et le processus recommence. La permutation des files consiste dans l'échange du contenu de la file principale et de la file secondaire. Les deux critères présentés précédemment agissent comme des contraintes garantissant, par définition, que le schéma avec double file de priorité n'entre jamais dans une boucle infinie (c'est-à-dire ils interdisent que les éléments se retrouvent à être échangés indéfiniment entre les deux files). Ce mécanisme est particulièrement efficace pour le gain de robustesse dans le cas d'occultations partielles.

Si la stratégie des files de priorité n'était pas utilisée, chaque calcul de la pose d'un élément serait fait indépendamment, et l'ensemble de ces calculs seraient fait de manière aléatoire. Afin de calculer le coût de l'utilisation de cette stratégie, nous devons considérer :

- n , le nombre d'éléments à traiter,
- t_p , le temps de calcul de la pose d'un élément,
- $ItMax$, le nombre maximum d'itérations dans la boucle,
- $s \in [0, n[$, le nombre d'éléments dont la pose peut être calculée à l'aide d'un opérateur,
- t_e , le temps d'évaluation de la pose d'un élément, c'est-à-dire le temps de calcul de $PoidPose3D_{(j,t)}$. $t_e \ll t_p$, parce que cette opération est réalisée plusieurs fois dans le calcul de la pose,
- t_o , le temps de calcul nécessaire pour l'application d'un opérateur afin de calculer la pose d'un élément. $t_o \ll t_p$, parce que cette opération nécessite un calcul simple (généralement, une ou deux multiplications de matrices),
- t_q , le temps de calcul pour extraire un élément dans une file de priorité ($t_q = O(\log(n))$).

Le temps de calcul total, sans l'utilisation de la stratégie des files de priorité, est égal à $n \times t_p$. Lorsque on utilise les files de priorité, le temps de calcul total a une limite maximum, à un facteur multiplicatif près, égal à :

$$(n \times (ItMax - 1) \times (t_q + t_e)) + (s \times t_o) + ((n - s) \times t_p) \quad (4.33)$$

Schématiquement, le temps de traitement global est asymptotiquement dominé par le facteur t_p , qu'on utilise ou pas les files de priorité. Cependant, la principale différence en temps de calcul est donnée par l'exploitation des dépendances au travers des opérateurs définis dans la description de la scène. Ce principe est un facteur important pour réduire les temps d'exécution. En effet, quand on utilise la stratégie des files de priorité, le gain est d'autant plus important que la valeur de s se rapproche de la valeur de n . Le modèle de la scène doit donc être construit en prenant ce principe en considération.

Chapitre 5

Reconnaissance d'événements et de scénarios

La vidéosurveillance automatisée (ou intelligente) est de plus en plus souvent intégrée comme outil de contrôle dans les environnements présentant des besoins exigeants en termes de sécurisation. Ce type de système s'appuie sur l'analyse de ce qui est « vu » dans les images vidéo, afin de détecter et de reconnaître des comportements ou des situations anormales dans les scènes filmées. Dans la plupart des cas, il s'agit de générer des alertes qui informent l'utilisateur et lui permettent d'intervenir pour corriger la situation, compenser les conséquences, ou dans le meilleur des cas, éviter toute situation générant des conséquences négatives à l'environnement surveillé et apporter des améliorations.

La manque d'efficacité de la vidéosurveillance « manuelle » provient notamment de la difficulté pour un humain de se concentrer constamment sur des moniteurs affichant des activités en continu. De plus, ces activités peuvent être interprétées différemment soit par un même individu visionnant la même activité à deux instants différents, soit par chaque individu regardant la même scène au même instant. Pour ces raisons, le traitement et l'analyse de données haut niveau issues de flux vidéo ont été un domaine de recherche très actif au cours des dernières années. Les outils utilisés pour la vidéosurveillance automatisée relèvent d'un domaine de recherche pluridisciplinaire, comme par exemple, le développement de nouveaux algorithmes de détection de mouvements pour systèmes embarqués [VVDM11], l'utilisation de l'analyse de la marche comme une donnée biométrique [GBCN10], ou encore l'intégration d'informations provenant d'autres moyens d'acquisition que la vidéo [MWWW11].

Un des principaux défis de ce type de système est qu'il doit s'exécuter de manière prévisible, la difficulté résidant dans le fait que les scénarios potentiels peuvent être de nature incertaine, bien que les conditions et les prérequis dans l'environnement soient bien définis. Dans ce cadre, les besoins de l'utilisateur définissent la performance attendue pour le système, comme par exemple, le taux de détections et celui de fausses alarmes. Assurer la performance du système tout au long de son exécution est également un autre défi à relever, en ayant souvent recours à des outils de mesure et de comparaison par rapport aux données de la « vérité terrain », à tous les niveaux : de l'acquisition des images jusqu'aux scénarios reconnus [BBE⁺08].

Afin de répondre à ces problématiques, il est parfois nécessaire d'établir un compromis et/ou de poser des hypothèses afin d'améliorer les résultats du système pour une application spécifique. Ainsi, par exemple, Velten et al. [VKK02] proposent un système pour la sécurisation de l'opérateur d'une machine industrielle spécifique. Ce système exploite des informations *a priori* et il a été conçu de manière à compléter les dispositifs habituels de sécurité, ces derniers n'étant pas suffisants pour garantir un bon fonctionnement en dehors des zones à risque. Dans le domaine de la sécurité publique, Chang et al. [CLC10] ont mis en place un système pour la détection de bagages abandonnés, par la détection des bagages et la recherche, l'identification et le suivi de leurs propriétaires dans l'entourage proche du bagage dans la scène. Dans le domaine du bâtiment, Gualdi et al. [GPC11] exploitent les informations contextuelles extraites des images d'un chantier filmé afin d'obtenir une meilleure robustesse dans la classification des objets. Enfin, lorsque le système de vidéosurveillance n'est pas restreint à un domaine étroit d'utilisation, comme ci-dessus, le compromis est donné au niveau de l'automatisation, le système opérant donc de manière supervisée.

Dans notre cadre, nous nous sommes principalement intéressés à l'analyse du flux opérationnel des tâches réalisées par le personnel médical pendant le déroulement du traitement en radiothérapie. Nous nous focalisons sur les tâches qui sont visibles à l'intérieur de la salle, le personnel étant présent ou non – par exemple, certaines machines peuvent être commandées depuis l'extérieur de la salle, et le déclenchement de l'irradiation, fait obligatoirement de l'extérieur, provoquera l'allumage d'un voyant lumineux dans la salle. Des travaux récents se sont déjà intéressés à la vidéosurveillance automatisée du flux de travail dans des environnements industriels, avec des besoins proches des nôtres [VGMVG11, MSR⁺10, VGK⁺10, NGG11]. Les caractéristiques de ces environnements rendent très complexe la tâche de surveillance, contrairement aux salles de traitement en radiothérapie dont les caractéristiques (cf. section 4.3) et l'accès aux informations *a priori* facilitent le traitement plus précis des informations 3D, qui sont un besoin fondamental en radiothérapie.

Dans le contexte médical, des travaux se sont également déjà intéressés à la modélisation du flux de travail dans le but d'assister la prise de décision [WKAS78, SBSR05]. Plus récemment, Malarme et al. [MWW11] ont proposé une méthode d'enregistrement manuel de « modèles de processus chirurgical », dans le but d'analyser leur déroulement par rapport à un modèle générique et de réduire les risques lors de l'opération chirurgicale. Heureusement, nous avons l'avantage de disposer d'une « feuille de route » du déroulement d'une séance de traitement en radiothérapie : le plan de traitement (issu du SPT). Des systèmes de surveillance de salles d'opération basés sur des modèles de Markov cachés ont également été présentés, en exploitant des informations 2D extraites à partir de plusieurs caméras [BJD11], ou bien des informations 3D acquises à partir d'un système de caméras [PMW⁺09]. Cependant, cette modélisation n'est pas adaptée à nos besoins car, dans notre cas, l'historique des événements passés conditionne la prédiction des événements futurs.

5.1 Gestion des erreurs en radiothérapie par vidéo

Les événements indésirables sont des incidents dans lesquels le traitement médical devient involontairement nocif pour la santé du patient. Ces événements proviennent

généralement d'une variété de caractéristiques systémiques qui se sont accumulées sur toutes les étapes du traitement. Bien que plusieurs systèmes de contrôle de qualité aient été mis en place, les erreurs et les effets indésirables sont encore fréquents dans la pratique clinique. Le facteur humain joue un rôle majeur dans la gestion des erreurs, mais il faut également prendre en compte le fait que les causes d'erreurs sont fortement liées au contexte et aux conditions de travail. Il y a donc une double approche lorsqu'on s'intéresse aux erreurs de chaque individu, car l'erreur commise dépend de son comportement (oubli, première expérience d'une erreur, relâchement d'attention, baisse de moral, etc.) et des facteurs préexistants qui génèrent les conditions dans lesquels l'erreur se produit (choix de gestion et d'organisation par les responsables, environnement stressant, systèmes de communication inadéquats, etc.) [Lea94].

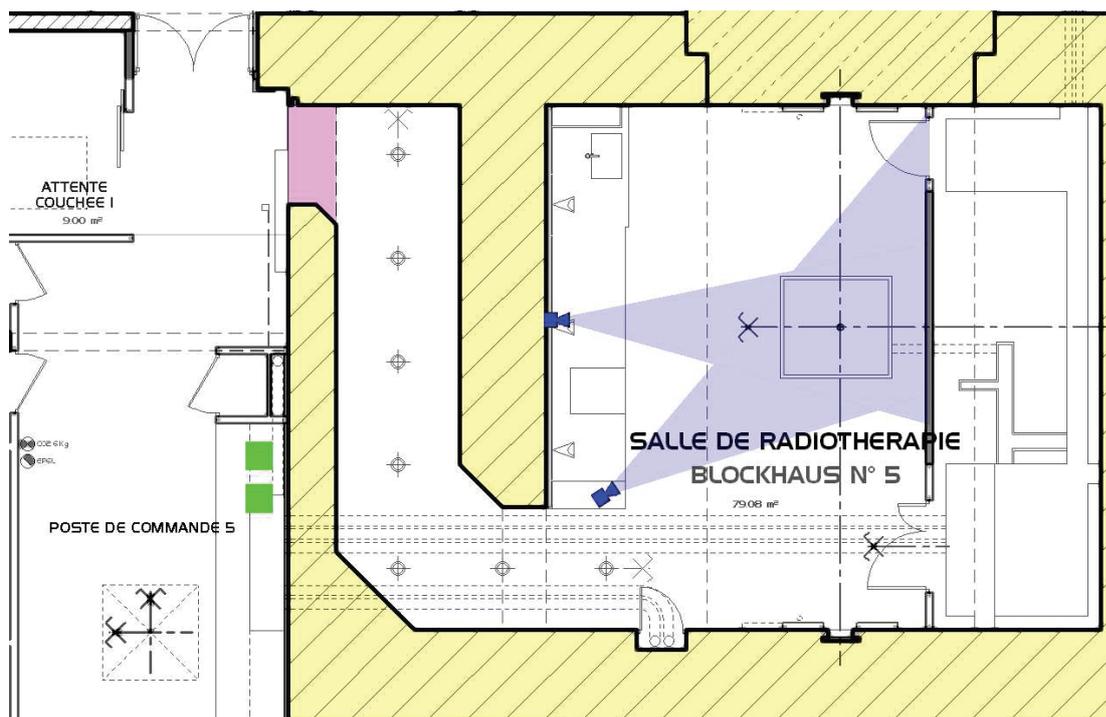
5.1.1 Utilisation de la vidéo dans une salle de traitement

Il est aujourd'hui d'usage d'avoir des caméras à l'intérieur de la salle de traitement, qui sont connectées à des écrans présents dans le poste de contrôle (ou poste de commande). À titre d'exemple, la Fig. 5.1 présente la configuration d'un tel système de vidéosurveillance (non automatisé). Sur cette configuration, le personnel médical est capable de voir l'intérieur de la salle depuis le poste de contrôle, notamment la zone où se situe le patient. Les images acquises ne sont pas traitées ni enregistrées, et elles servent uniquement de témoin visuel lors du déroulement de la séance. Un interphone permet de communiquer entre l'intérieur de la salle et le poste de contrôle, ce qui permet au personnel de donner des consignes aux personnes présentes dans la salle depuis l'extérieur, et au patient d'entendre leurs réactions et de faire part au personnel d'un problème particulier.

Bien que la Fig. 5.1 ne montre qu'un exemple, la configuration générale des salles de traitement en radiothérapie est assez proche. Comme on peut le voir, les personnes qui entrent ou sortent de la salle doivent traverser un couloir. Ce couloir est un moyen de radioprotection qui doit être utilisé pour se protéger lors qu'on reste enfermé (par accident) à l'intérieur de la salle. En effet, le mur permet de réduire l'exposition à l'irradiation, et un bouton d'alarme permet immédiatement d'alerter le personnel présent à l'extérieur pour que l'irradiation soit arrêtée. Cependant, l'existence de ce couloir est l'un des principaux défauts des systèmes de vidéosurveillance actuels, car le personnel ne voit plus l'intérieur de la salle lorsqu'il traverse ce couloir. Un autre défaut est que le personnel, lors qu'il se trouve à l'extérieur de la salle, doit constamment regarder les écrans afin de détecter / déceler / anticiper / contrôler toute situation dangereuse. Même dans le cas où le personnel est capable d'observer à tout instant l'intérieur de la salle (plus d'une personne est présente pour délivrer le traitement), les différentes tâches qu'il doit réaliser font qu'il est difficile de se concentrer en continu sur les écrans reliés aux caméras et sur le contrôle de situations interdites ou dangereuses pour le bon déroulement du traitement.

5.1.2 Vidéosurveillance automatisée d'une salle de traitement

L'utilisation de techniques de vision par ordinateur et d'intelligence artificielle permet de combler les manques des systèmes de vidéosurveillance actuels, à tra-



(a) Configuration de la salle. Bleu : caméras et champ de vision. Vert : écrans connectés aux caméras. Violet : seule entrée de la salle. Jaune : murs blindés.



(b) Caméra et écran 1

(c) Caméra et écran 2

FIGURE 5.1: Exemple de configuration d'un système de vidéo surveillance dans salle de traitement en radiothérapie (source : Centre Léon Bérard [7]).

vers l'apport d'informations sur la quantification des paramètres des éléments de la salle, et à travers l'assistance dans le contrôle du déroulement du traitement. Plus précisément, dans notre cas, le suivi de mouvement des éléments dans la salle (tel que présenté dans le chapitre précédent) permet de connaître la pose 3D de chaque élément dans la scène. Ces informations sont transmises au personnel et comparées aux informations contenues dans le plan de traitement, ce dernier contenant les poses planifiées. L'analyse de la configuration courante, définie comme un ensemble de valeurs de paramètres des éléments présents dans la salle, permet de contrôler que cette configuration est conforme à celles planifiées et aux consignes de sécurité. Additionnellement, l'analyse temporelle des configurations permet de contrôler le bon déroulement de la séance, et notamment que toutes les configurations planifiées soient « visitées » comme prévu. On doit comprendre par « configuration visitée », une configuration prédéfinie pour laquelle, à un instant donné, la configuration courante de la scène a été conforme. Ces deux derniers points, autour de l'analyse des

configurations, sont présentés dans la suite de ce chapitre, en utilisant les notions suivantes :

un élément est un individu (patient ou personnel) ou un objet (machine, équipement, etc.) potentiellement présent dans une salle de traitement, avec des valeurs de paramètres bien définies,

un état d'un élément est un sous-espace de *EVP*, l'espace de valeurs possibles des paramètres d'un élément, et l'union de tous les états est égale à *EVP*. De manière analogue, on définit **un état d'un ou plusieurs paramètres d'un élément**. L'état d'un élément permet de savoir notamment si l'élément est conforme au plan de traitement et aux consignes de sécurité, alors que l'état d'un ou plusieurs paramètres de cet événement permet d'évaluer l'état de l'élément (par exemple, les paramètres de la pose d'un élément a deux états : bonne position ou mauvaise position, alors que l'identification de l'élément est un paramètre à prendre également en compte, entre autres).

une configuration est un couple de vecteurs : un vecteur d'états (un pour chaque élément) et un vecteur de valeurs de paramètres (pour tous les éléments de la scène – par exemple, les paramètres de la pose 3D), les valeurs de ces deux vecteurs étant liées. Lors qu'il s'agit d'une configuration prédéfinie, les vecteurs ont des valeurs fixes. On définit une configuration conforme et visitée comme suit :

- c_0 est une **configuration conforme** à une configuration c_1 si les valeurs des composants des deux vecteurs sont égales à un epsilon près (précision de l'acquisition des données).
- une configuration prédéfinie c est une **configuration visitée** à un instant t , si elle est conforme à la configuration courante à l'instant t , ou si elle a été conforme à la configuration courante à un instant $t' < t$.

un événement est, pour un élément, le changement de son état par rapport à l'instant précédent (changement instantané),

un scénario est une suite d'événements ayant des liens temporels entre eux. Il décrit le comportement d'un système dynamique pendant une durée.

5.2 Modélisation d'une séance de traitement

Dans le but de bien identifier les paramètres externes de l'administration du traitement, notamment pour les éléments et leurs interactions, nous avons créé des schémas basés sur les états des éléments lors du déroulement d'une séance de traitement (cf. section 2.1.3). Cette modélisation a été réalisée à l'aide de radiophysiciens du Centre Léon Bérard [7]. Le début d'une séance est marqué par le choix d'un plan de traitement. Cela se traduit par le fait que, depuis le début de la séance, on sait à tout instant quels éléments doivent être présents, principalement le patient (c'est-à-dire, son identité), et quelles configurations doivent être visitées.

Les schémas utilisés sont basés sur des arbres de décision contenant des éléments et des événements potentiellement présents dans la scène. Les différents arbres sont destinés à n'en former qu'un seul¹, cet arbre contenant au niveau des feuilles des

1. le document contenant tous les arbres personnalisés n'est pas inclus dans ce manuscrit parce qu'il est non diffusable.

zones de danger dans lesquelles peut se trouver le système. Ces schémas pourront, à l'avenir, servir à définir un modèle du système des séances de traitement en radiothérapie. Suivant la représentation donnée dans la Fig. 5.2, la racine d'un arbre (couleur cyan) est associée à une feuille dans un autre (couleur magenta), à l'exception du nœud « scène » qui est la racine globale.



FIGURE 5.2: Continuité dans les schémas.

Un premier arbre (Fig. 5.3) contient les différents type d'éléments, classifiés selon leur nature (regroupement par sous-arbre, et classe définie par le nœud racine du sous-arbre). Dans cet arbre, les nœuds « Élément X » (couleur magenta) représentent un identifiant associé au type d'élément contenu dans le nœud père.

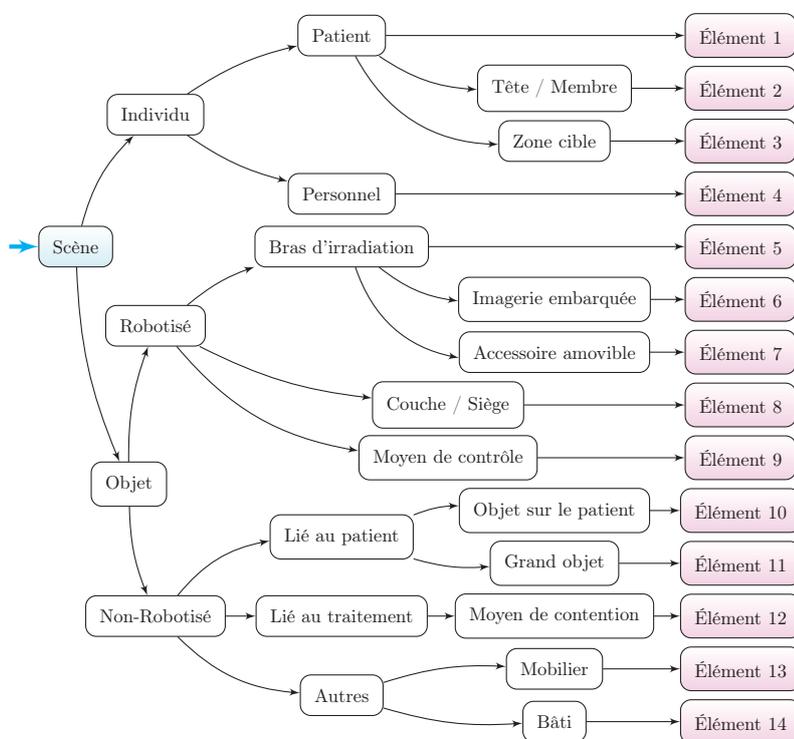


FIGURE 5.3: Éléments potentiellement présents dans la scène.

L'arbre de la Fig. 5.4 est une représentation générique des événements potentiels de chaque type d'élément. C'est notamment à ce niveau que les caractéristiques globales de la scène sont prises en compte. La tendance actuelle dans la pratique clinique est de ne rien tolérer pendant l'irradiation, c'est-à-dire, lorsque l'accélérateur est en marche, et de contrôler que des événements considérés comme dangereux n'arrivent pas hors irradiation. C'est pour cette raison que le facteur « irradiation » est l'un des premiers pris en compte (dans le parcours d'un chemin de l'arbre).

On cherche ensuite à déterminer si la présence de l'élément est autorisée (nécessaire ou tolérée) ou interdite, selon ce qui a été planifié et selon les consignes de sécurité. L'exemple le plus parlant est : on tolère la présence du patient et du personnel (qui intervient pour positionner le patient, mettre en place ou retirer un

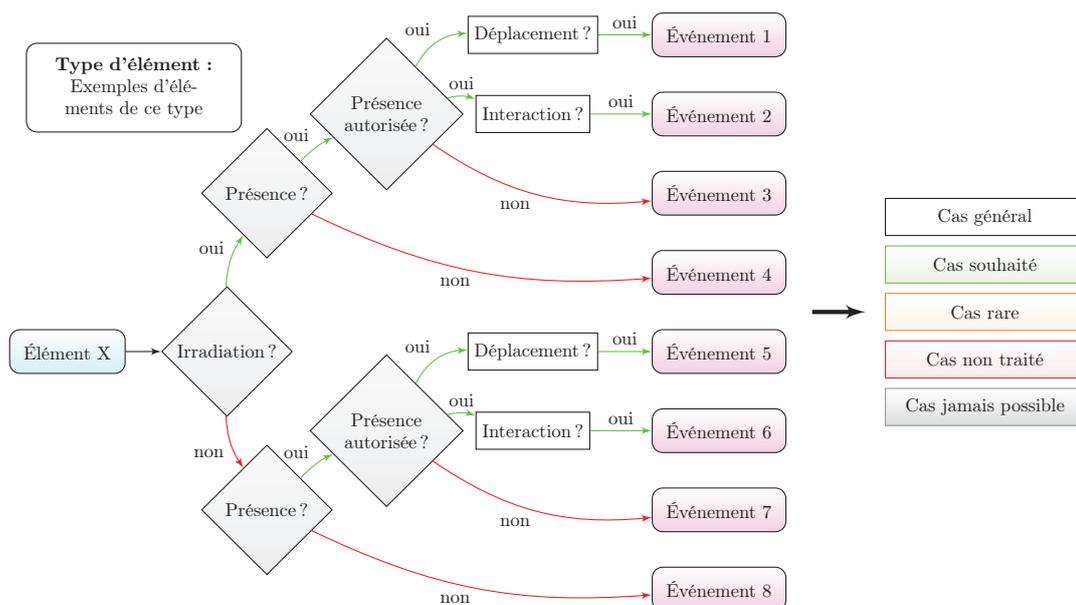


FIGURE 5.4: Événements potentiels.

accessoire, etc.) hors irradiation, et pendant l'irradiation la présence du patient est nécessaire mais la présence du personnel est interdite. Un exemple d'un élément autorisé mais pas nécessaire, est le cas d'une chaise ou d'un tabouret, qui n'est pas indispensable pour délivrer le traitement mais qui peut aider le patient (et même le personnel) à être plus confortable : le patient peut s'asseoir pendant que le personnel prépare la table de traitement, etc..

Deux autres critères sont ensuite considérés dans l'arbre de la Fig. 5.4. Le premier est la position de l'élément, qui n'est évaluée que lorsque l'élément se déplace, et le deuxième porte sur les interactions potentielles avec les autres éléments dans son entourage proche. Pour ce dernier, un espace d'interaction est associé à chaque élément. Cet espace est représenté par un volume défini par la géométrie de l'élément et l'ensemble de poses dans lesquelles l'élément peut être présent pendant une durée de temps fixe et prédéfinie, à partir de la pose courante. Si les espaces d'interaction de deux éléments s'intersectent, on détecte alors :

- une interaction potentielle, si au moins un de deux éléments est un individu, ou si au moins un de deux éléments est en même temps en interaction avec un individu,
- une collision potentielle dans tous les autres cas.

Les événements contenus dans les feuilles ont une classe associée² qui indique la probabilité que l'événement ait lieu :

1. cas général : peut arriver dans toute circonstance,
2. cas souhaité : ce qu'on cherche à avoir (le bon patient, à la bonne position, etc.),
3. cas rare : peut arriver exceptionnellement (le patient tombe de la table, panne de courant, etc.),
4. cas non traité : par l'aspect générique des arbres, les événements relatifs à certains éléments sont traités dans les événements d'un autre élément (la présence

2. valeurs personnalisées dans l'arbre global, et définies par les experts.

du la zone à irradier est déjà traitée lors de la présence du patient, bien qu'on différencie le comportement de ces deux éléments),

5. cas jamais possible : également par l'aspect générique des arbres, certains événements ne peuvent pas arriver pour un élément (par exemple, l'absence du bras d'irradiation est impossible).

Trois arbres génériques complètent la structure globale (Fig. 5.5, Fig. 5.6 et Fig. 5.7). Il s'agit de classifier l'événement, en fonction des paramètres de l'élément et des informations de planification, afin de définir si l'élément est :

1. dangereux (zone de danger),
2. non dangereux (ou tolérable) mais avec un risque potentiel (pour une durée de temps courte, fixe et prédéfinie – zone de transition),
3. non dangereux et sans risque potentiel (zone sans danger).

Trois informations sont associées à chacune de ces zones : la description sémantique de l'événement, l'interaction qui doit être réalisée entre le module de contrôle et le personnel (action), et la conséquence que l'événement peut avoir sur le système, c'est-à-dire, sur les autres éléments et ses interactions (pas sur la santé du patient).

La classification est faite par la confrontation des valeurs des paramètres de l'élément et des valeurs de référence (une illustration des résultats est donnée à la Fig. 5.8), ces derniers pouvant être variables dans le temps (plusieurs configurations prédéfinies). Par exemple, la position du patient est confrontée à la position cible (position unique dans le cas général de traitements isocentriques, c'est-à-dire, le volume à irradier à l'isocentre, point traversé par tous les faisceaux), ou bien, la position du bras d'irradiation est confrontée aux positions prédéfinies pour les faisceaux. La description précise de la manière dont les événements sont classifiés est donnée à la section 5.3.

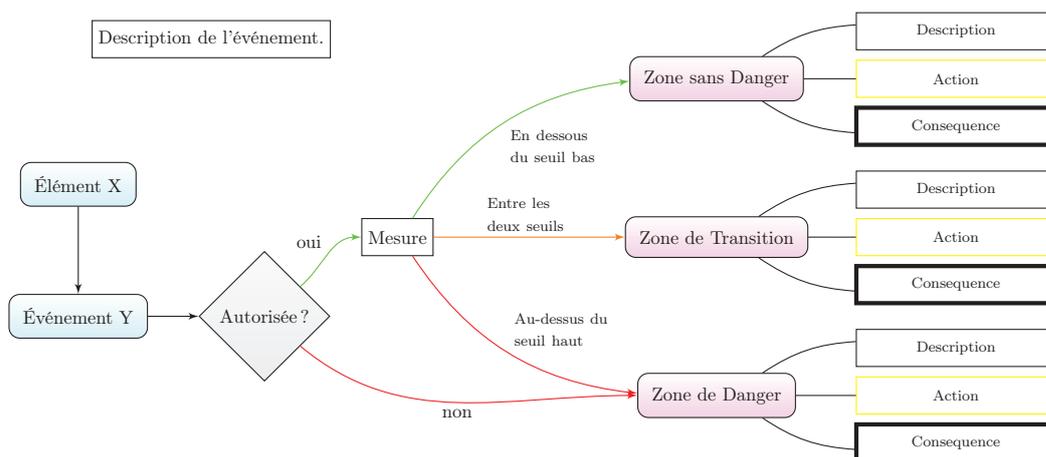


FIGURE 5.5: Événement de type déplacement.

L'arbre de la Fig. 5.5 est associé aux événements « déplacement ? » (événements 1 et 5 dans la Fig. 5.4). Un facteur supplémentaire est pris en compte : l'autorisation du mouvement. Cela est généralement associé aux consignes de sécurité, bien qu'on s'attache également à vérifier que les éléments restent conformes à une configuration prédéfinie pendant une durée de temps (lorsqu'on irradie notamment). Ainsi,

des mouvements autorisés peuvent devenir dangereux s'ils dépassent les marges de tolérance (seuils).

L'arbre de la Fig. 5.6 est associé aux événements « interaction ? » (événements 2 et 6 dans la Fig. 5.4). Dans ce cas, on aura autant d'événements de ce type qu'on a d'interactions potentielles avec d'autres éléments dans la scène (détectées à l'aide des espaces d'interaction comme expliqué précédemment). En plus de l'autorisation de l'interaction, on souhaite déterminer si l'interaction est nécessaire pour le bon déroulement de la séance. Par exemple, il est nécessaire que le personnel entre dans la salle pour (re)positionner le patient, et il doit manipuler la table de traitement et assister le patient, autrement dit, des interactions entre le patient, le personnel et la table sont nécessaires avant le début de l'irradiation.

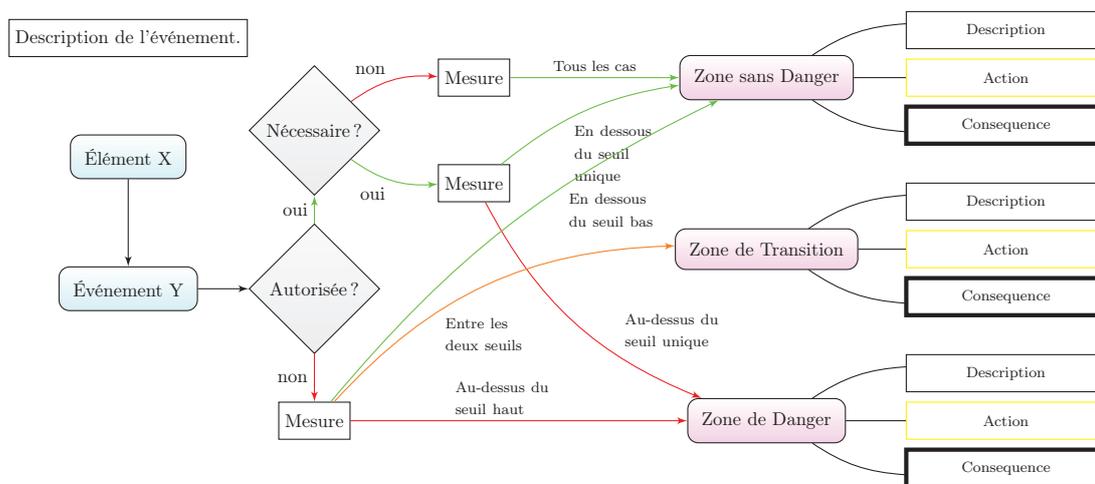


FIGURE 5.6: Événement de type interaction.

Le dernier arbre, présenté dans la Fig. 5.7, est associé aux événements liés à la présence de l'élément (événements 3, 4, 7 et 8 dans la Fig. 5.4). Dans ce cas, la zone associée à l'événement est unique (valeur personnalisée dans l'arbre global).

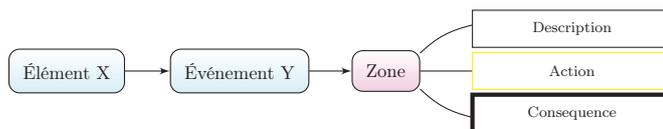


FIGURE 5.7: Autres type d'événement.

5.3 Reconnaissance d'événements

La détection et la reconnaissance d'événements est réalisée en parcourant l'arbre présenté précédemment, à chaque instant t et après avoir réalisé le suivi de mouvement des éléments et la mise à jour des variables globales (irradiation, identifiant du patient attendu selon le plan d'irradiation courant, etc.). Pour chaque élément d'intérêt³, dont l'un des paramètres a varié par rapport à l'instant précédent, nous

3. ceux contenus dans la file de priorité présentée à la section 4.5.4.

parcourons tous les chemins possibles dans l'arbre à partir de la racine en fonction de la configuration actuelle de la salle (variables globales). À partir de chaque chemin parcouru depuis la racine jusqu'à l'une des feuilles, on considère qu'un événement a été reconnu, et on extrait un triplet (*élément*, *événement*, *état*), que nous appellerons triplet *incident*, où *état* est l'état de l'élément (sans danger, transition ou danger). Des triplets de ce type peuvent également être générés par les autres modules, principalement l'interface graphique.

5.3.1 Classification d'un événement

La classification d'un événement permet d'obtenir l'état d'un élément par rapport à un événement détecté, et elle est faite en comparant les valeurs de paramètres de l'élément aux valeurs de référence. La différence entre les valeurs courantes des paramètres de l'élément et les valeurs de référence est comparée à deux seuils fixes et prédéfinis : un seuil haut et un seuil bas (action effectuée dans les nœuds « mesure » des arbres Fig. 5.5 et Fig. 5.6). Plus précisément, $EVDP$, l'espace des valeurs possibles de la valeur absolue de cette différence, est partitionné⁴ en trois sous-ensembles, un pour chaque classe : sans danger, transition et danger. Pour leur construction, les valeurs possibles de chaque paramètre sont divisées en trois intervalles. Chaque sous-ensemble est donc constitué d'un ensemble d'intervalles de valeurs, un intervalle pour chaque paramètre, et les valeurs limites entre les intervalles permet de définir deux vecteurs seuil :

- le seuil bas est défini par un vecteur contenant les valeurs limites entre les sous-ensembles sans danger et transition,
- le seuil haut est défini par un vecteur contenant les valeurs limites entre les sous-ensembles transition et danger.

Le seuillage est fait comme suit : soit V_p un vecteur contenant les valeurs courantes des paramètres de l'élément à considérer, et V_r un vecteur contenant les valeurs de référence, on calcule V_d , qui est le vecteur contenant la valeur absolue de la différence entre les valeurs de V_p et de V_r pour chaque paramètre. V_d est, par rapport à S (vecteur seuil), défini comme :

- en dessous du seuil : si les valeurs absolues de tous ses composants du vecteur sont strictement inférieures à la valeur de leurs homologues respectifs dans S ,
- au-dessus du seuil : sinon.

Les seuils bas sont toujours inférieurs aux seuils hauts. Sur cette base, un événement est classé dans la zone :

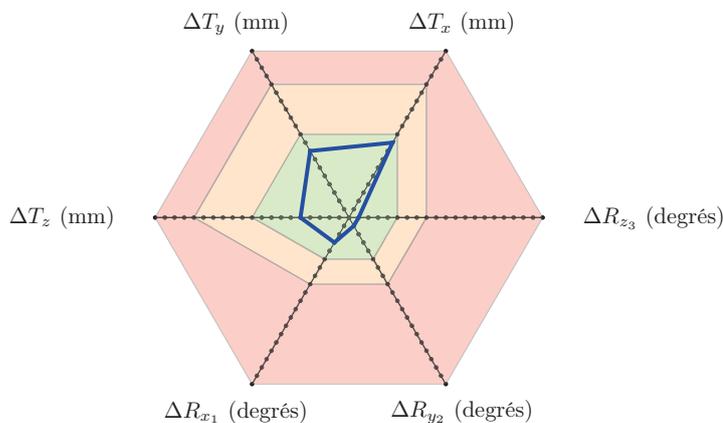
- sans danger, si les valeurs V_d sont en dessous du seuil bas,
- de transition, si les valeurs V_d sont au-dessus du seuil bas et en dessous du seuil haut,
- de danger, si les valeurs V_d sont en dessus du seuil haut.

Cette classification peut être vue autrement. Si l'on considère un polygone « configuration » p (illustré par un trait bleu à la Fig. 5.8) qui contient les valeurs courantes des paramètres de l'élément, l'événement est classé dans la zone :

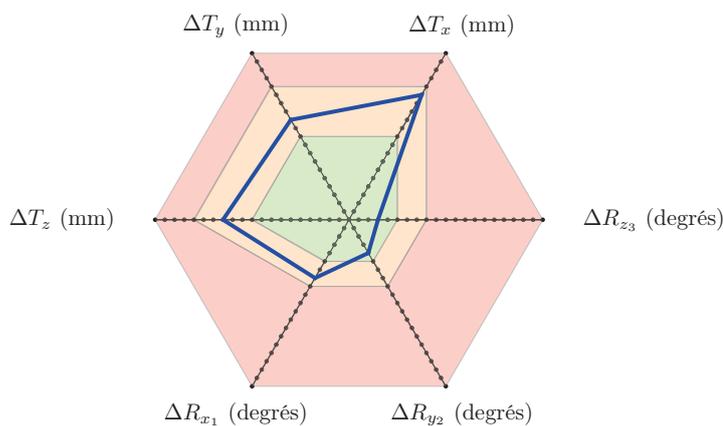
- sans danger, si p est inclus entièrement dans le sous-ensemble sans danger,

4. Des sous-ensembles de E forment une partition de E si et seulement si leur union est égale à E et leur intersection, deux à deux, est égale à l'ensemble vide.

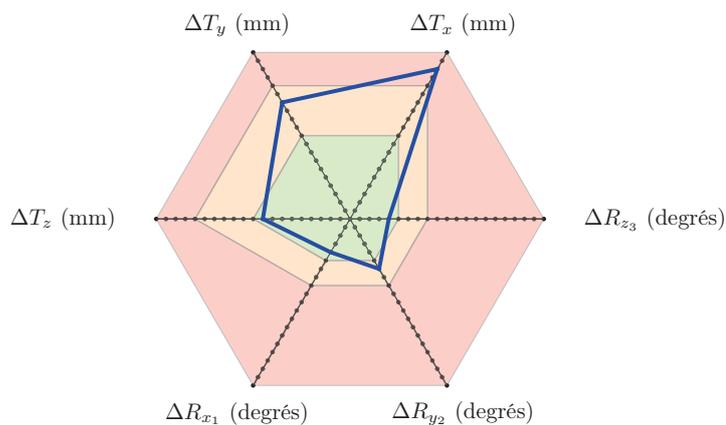
- de transition, si p est inclus dans l'union des sous-ensembles sans danger et transition, et p a une intersection non vide avec le sous-ensemble transition,
- de danger, si p est inclus dans l'union des sous-ensembles sans danger, transition et danger, et p a une intersection non vide avec le sous-ensemble transition danger.



(a) Configuration en état sans danger



(b) Configuration en état de transition



(c) Configuration en état de danger

FIGURE 5.8: Exemple de diagrammes d'états d'un élément portant sur les paramètres de la pose uniquement (une configuration avec un état différent dans chaque sous-figure, et les différentes valeurs de paramètres sont reliées par un trait bleu). Sous-ensembles : sans danger (région verte), transition (région orange) et danger (région rouge).

Dans la pratique, les valeurs de seuils sont définies par les experts dans le domaine de la radiothérapie, en tenant compte de la précision dans l'acquisition des données.

5.3.2 Analyse des événements

Au niveau des événements, il est déjà possible d'extraire des informations permettant d'alerter le personnel sur des situations dangereuses. Il s'agit notamment de traiter, selon leurs états, les triplets *incident* extraits. Nous avons proposé et développée une première approche basée sur des points de contrôle. Cette approche analyse le flux d'événements, sans s'intéresser à l'aspect temporel, et vérifie, à la demande expresse de l'utilisateur, la conformité entre la configuration courante et les configurations prédéfinies pour les faisceaux.

Nous introduisons le principe de « séance valide »⁵ (dans le cadre de la radiothérapie conventionnelle), pour le plan de traitement courant, comme une séance où :

1. le bon patient a été irradié au bon endroit selon les N faisceaux.
2. personne d'autre que le bon patient n'a été irradié (c'est-à-dire, une seule personne est à l'intérieur de la salle lorsqu'on irradie, et cette personne a été bien identifiée comme étant le patient attendu).
3. chacun des N faisceaux n'est administré qu'une et une seule fois.
4. chacun des N faisceaux correspond à une configuration valide de l'ensemble des éléments.
5. aucun événement n'est détecté durant l'irradiation (faisceau actif).

Sur ce principe, les seuils pour la reconnaissance d'événements sont définis de manière que le système puisse avoir une grande souplesse hors irradiation et être très strict pendant l'irradiation. Les triplets *incident(s)* où l'état est *transition* ou *danger* génèrent des alertes pour l'utilisateur. On convient que les triplets *incident* avec un état *transition* peuvent être corrigés par un autre triplet *incident* spécifique et présentant un état *sans danger*, et les triplets *incident* ayant un état *danger* demandent l'arrêt instantané de l'administration du traitement (irradiation principalement).

Une deuxième tâche importante est la vérification des configurations prédéfinies. Pour cela, on ajoute au système un événement provenant de l'interface utilisateur. C'est l'événement « pression sur le bouton irradiation » associé à l'interface utilisateur et qui permet au paramètre « demande irradiation » (de l'interface) de basculer entre les états « démarrage » et « arrêt » (le paramètre étant initialement sur l'état « arrêt »). Quand l'utilisateur appuie sur le « bouton irradiation » de l'interface, la valeur de la variable « demande irradiation » change et l'interface génère un triplet *incident* :

1. (*interface utilisateur, pression sur le bouton irradiation, transition*) quand le paramètre « demande irradiation » passe de « arrêt » à « démarrage »,
2. (*interface utilisateur, pression sur le bouton irradiation, sans danger*) dans le cas contraire.

5. la notion de validité est donné par rapport aux informations externes uniquement.

Algorithme 2 Algorithme de vérification de la validité de la séance.

ENTRÉES:

- $E(ELEM, EVENT, ETAT)$: triplet incident à traiter,
- IRR : **vrai** si on irradie, **faux** sinon,
- ES : état de la séance (succès, autorisation, conforme, attention ou danger),
- L : liste de triplets (*élément, événement, état*),
- C : ensemble de configurations prédéfinies,
- c : configuration actuelle de la salle.

SORTIES: mise à jour (si nécessaire) de ES

si $ETAT = danger$ **alors**

$ES \leftarrow danger$

sinon

si non IRR **et** $E = (interface\ utilisateur, pression\ sur\ le\ bouton\ irradiation, transition)$ **alors**

on recherche la configuration $c' \in C$ la « plus proche » de c

si c conforme à c' **et** c' non visitée **alors**

$ES \leftarrow autorisation$

sinon

$ES \leftarrow attention$

fin si

sinon

si $ETAT = transition$ **alors**

$ES \leftarrow attention$

ajouter E dans L

sinon

si non IRR **et** $E = (interface\ utilisateur, pression\ sur\ le\ bouton\ irradiation, sans\ danger)$ **et** toutes les configurations ont été visitées **alors**

$ES \leftarrow succès$

sinon

si $ETAT = sans\ danger$ **et** E corrige un triplet E' dans L **alors**

enlever E' de L

si L est vide **alors**

$ES \leftarrow conforme$

fin si

fin si

fin si

fin si

fin si

fin si

Un module d'analyse d'événements a été mis en place, constitué notamment des opérations contenues dans l'algorithme 2. Lorsque le triplet (*interface utilisateur, pression sur le bouton irradiation, transition*) est reçu par le module, on vérifie si la configuration courante est conforme à une configuration de faisceau non visitée. Si c'est le cas, l'irradiation est autorisée et commence : on génère le triplet (*interface utilisateur, autorisation irradiation, sans danger*), qui doit créer un signal visuel (interface graphique) pour indiquer au manipulateur qu'il est autorisé à mettre en

marche l'irradiation. Sinon, on génère le triplet (*interface utilisateur, autorisation irradiation, transition*) qui doit créer un signal visuel indiquant à l'utilisateur les paramètres (dans la configuration courante) à corriger pour atteindre une configuration prédéfinie non visitée (la « plus proche », par exemple, en utilisant une métrique sur certains paramètres).

Un état global pour la salle *ES* est défini, et modifié en fonction du flux d'événements. Selon la valeur de cet état, plusieurs types d'informations sont données à l'utilisateur :

l'état succès signifie que tous les configurations prédéfinies ont été visitées, et que le personnel peut entrer dans la salle pour aider le patient à en sortir,

l'état autorisation indique que la configuration courante est prête pour l'un des faisceaux et que l'irradiation peut commencer,

l'état conforme signale que tout va bien dans la séance,

l'état attention indique qu'il y a un ou plusieurs paramètres dans la scène qui doivent être corrigés,

l'état danger alerte que l'administration du traitement doit être arrêtée immédiatement (arrêt de l'irradiation notamment) à cause d'une erreur qui ne peut pas être corrigée.

On différencie donc les triplets *incident* qui peuvent être corrigés de ceux qui ne le peuvent pas⁶. En effet, si le personnel est capable de corriger une erreur, la séance sera poursuivie et achevée avec succès. La séance terminée (le patient est à l'extérieur de la salle et on est prêt à changer de plan de traitement pour le prochain patient), cette séance sera considérée comme validée, d'après les critères présentés précédemment, si l'état global de la salle *ES* a pris la valeur *succès* une fois pendant la séance et *ES* n'a pris aucune fois la valeur *danger*.

Cette approche a l'avantage d'être simple à mettre en place et de s'exécuter en quelques opérations. Elle permet principalement d'évaluer la conformité du déroulement de la séance par rapport aux données « a priori » et de détecter des erreurs quand elles arrivent, mais elle souffre d'un manque total d'analyse temporelle des événements. Cette analyse temporelle permettrait de prédire les événements qui vont arriver, voire de détecter des erreurs avant qu'elles ne se produisent. Nous proposons une deuxième approche, basée sur la reconnaissance de scénarios à partir d'un flux d'événements, qui peut répondre à cette problématique.

5.4 Reconnaissance de scénarios

La reconnaissance de scénarios est un sujet de recherche très actif dans le domaine de l'intelligence artificielle [DGG93, VBT03, YB03, RBT10]. Cette technique répond à un réel besoin d'automatisation de l'extraction d'informations sur les événements qui ont eu lieu, et peut fournir un diagnostic du système et « prédire » les événements potentiels à venir. La comparaison des informations sur les événements par rapport à des modèles de scénario permet de reconnaître des situations typiques ou atypiques,

6. au niveaux médical, les triplets *incident* ne pouvant pas être corrigés peuvent parfois être compensés dans la suite du traitement, cela en fonction de leur niveau de gravité.

même avant qu'elles aient pu se produire. Les modèles de scénarios peuvent être le résultat d'un algorithme d'apprentissage ou être des modèles prédéfinis construits à partir d'informations *a priori* [LBP⁺07].

Plusieurs travaux se sont déjà intéressés à la reconnaissance de scénarios comme outil permettant de compléter un système de vidéosurveillance [CGIP01, ELMM05]. Dans le domaine médical [DRF98, FGI01], la reconnaissance de scénarios est perçue comme un outil d'aide à la prise de décision, par exemple, en suivant l'état du patient et en proposant des actions à effectuer en accord avec le traitement, pour que ce dernier puisse être mené à terme et contribuer à son amélioration. Cette technique est aussi utilisée pour attester du bon déroulement des étapes dans un protocole médical.

5.4.1 Application à la vidéosurveillance de la salle

Pour les traitements en radiothérapie, le personnel médical dispose d'instructions à réaliser pour l'administration du traitement (plan de traitement). Comme ce type de traitement est hautement contrôlé et fortement contraint, il est possible de formaliser la majorité des scénarios possibles sur la base de ces instructions, avec l'aide d'experts. On est même capable de formaliser tout scénario correspondant à un déroulement conforme de la séance, avec la présence d'erreurs corrigées ou pas, car les instructions de l'administration du traitement sont précises. Tout scénario non reconnu représente forcément une situation inattendue, et donc dangereuse (sauf quelques cas particuliers n'étant pas été pris en compte lors de la formalisation des scénarios) pour le bon déroulement de la séance, voire la santé du patient. Sur ce principe, il n'y a pas besoin d'utiliser une méthode d'apprentissage pour créer de modèles de scénarios, et nous utiliserons de modèles définis *a priori*.

Une caractéristique supplémentaire de celles présentées précédemment pour une séance valide, doit être prise en compte. L'ordre temporel relatif de certaines instructions n'a pas d'importance, bien qu'elles soient exécutées entre deux événements bien définis. Par exemple, en radiothérapie conventionnelle, les faisceaux peuvent être administrés dans n'importe quel ordre, éventuellement différent pour chaque séance. L'important est qu'ils soient tous bien administrés, et une seule fois chacun.

Nous avons décidé d'utiliser une représentation par automate fini [BM98], plutôt que par des réseaux bayésiens [HBN00b] et par des graphes [VBT03]. Ce choix est justifié par le fait qu'il n'est pas nécessaire de modéliser toutes les combinaisons possibles de tous les événements, notamment pour des séquences d'événements devant se répéter (par exemple, le personnel entre plusieurs fois dans la salle pour corriger la même erreur), et que les relations temporelles sont plus simples à modéliser. Afin de donner de la souplesse, les états de l'automate sont régis par les configurations possibles de la salle de traitement, les événements agissant comme des modificateurs des états contenus dans les configurations. Ainsi, nous redéfinissons l'état *autorisation* et nous ajoutons un nouvel état *irradiation*, pour les valeurs de l'état global de la séance afin de prendre en compte l'aspect temporel. Soit C l'ensemble des configurations qui doivent être visitées, pour chaque configuration $c_j \in C$ on définit :

l'état *autorisation* c_j , qui indique que la configuration courante est prête pour l'administration du faisceau ayant la configuration c_j , et que l'irradiation peut commencer,

l'état *irradiation* c_j , qui indique que la configuration courante est conforme à la configuration c_j et que l'irradiation est en marche.

5.4.2 Modèle de scénario

Les états de l'automate définissent les états possibles pour la configuration courante de la salle, et les transitions sont régies par les événements et par le résultat de l'évaluation de la configuration courante de la salle. Cela revient à traiter les triplets *incident* en fonction de l'état global courant de la salle, et à modifier cet état si nécessaire. Pour des questions de recouvrement de termes, un état de l'automate sera également appelé *nœud* dans la suite de cette section.

Pour chaque état global de la salle, on définit deux ensembles EAC_e et ESC_e contenant, le premier, les triplets *incident* pouvant être corrigés, et le deuxième, les triplets *incident* ne pouvant pas l'être. Un nœud q_i est composé :

- d'un état global de la salle,
- d'un ensemble de transitions par triplet *incident* associé à ce nœud de l'automate. Nous appellerons AE_i l'ensemble de ces triplets,
- d'un ensemble de transitions par événement contenu dans $EAC_e \setminus (AE_i \cup ENR)$ (la différence des ensembles EAC_e et $AE_i \cup ENR$, où ENR est l'ensemble de triplet *incident* qui ne sont pas réalisables à partir de ce nœud), toutes en direction d'un nœud contenant l'état *attention*,
- d'un ensemble de transitions par événement contenu dans $ESC_e \setminus (AE_i \cup ENR)$, toutes en direction d'un nœud contenant l'état *danger*.

Plus clairement, on considère par défaut que tous les événements sont dangereux ou potentiellement dangereux. Pour chaque nœud, on définit donc les triplets *incident* ne présentant pas de danger pour ce nœud, afin de construire AE_i .

À toute transition, on peut associer une opération pouvant générer un triplet *incident* qui sera ensuite traité. Cette opération est exécutée après la transition, c'est-à-dire que le nœud courant dans l'automate lors de l'exécution de l'opération sera le nœud destination de la transition à laquelle l'opération est associée.

L'opération *évaluation*(c) a un statut spécial. Elle est utilisée pour l'évaluation de la conformité de la configuration courante c par rapport à l'une des configuration prédéfinies. Cette évaluation est faite à l'aide de l'utilisation de la mémoire (variables) dans l'automate, afin vérifier si une configuration prédéfinie a déjà été visitée et si le faisceau associé a déjà été administré. Dans cette opération, on cherche la configuration $c_j \in C$ la plus proche de la configuration courante (au sens d'une métrique sous les paramètres), C étant l'ensemble des configurations prédéfinies. La configuration c_j peut avoir l'un des trois états (dans la mémoire de l'automate) :

validée indique que le faisceau associé à c_j a été déjà administré, c'est-à-dire, qu'on a visité la configuration c_j et qu'on a irradié le patient dans cette configuration,

activée indique que la configuration courante est conforme à la configuration c_j , mais que le faisceau associé n'a pas encore été administré,

désactivée indique que la configuration courante n'est pas conforme à la configuration c_j et que le faisceau associé n'a pas encore été administré.

Lorsque l'opération *évaluation*(c) est exécutée, on a trois cas possibles. Soit $c_j \in C$ la configuration la plus proche de la configuration courante c :

1. si c_j est *désactivée*, et que c est conforme à c_j , on génère un triplet (*salle, identification c_j , sans danger*). Par construction, le nœud courant de l'automate (lors de l'exécution de l'opération) doit avoir une transition associée à ce triplet *incident*. La destination de cette transition, dans l'automate, est un nœud ayant l'état *autorisation c_j* . De plus, cette transition a une opération associée changeant l'état de c_j , dans la mémoire de l'automate, à l'état *activée*,
2. si c_j est *activée*, et que c n'est pas conforme à c_j , on génère un triplet (*salle, identification c_j , transition*). Par construction, le nœud courant de l'automate doit avoir une transition associée à ce triplet *incident*. La destination de cette transition est un nœud contenant l'état *attention*. De plus, cette transition a obligatoirement une opération associée changeant l'état de c_j à l'état *désactivée*,
3. pour toute autre combinaison des états de c_j et de la conformité de c par rapport à c_j (c n'est conforme à aucune configurations prédéfinie et aucune configuration prédéfinie n'est *activée*, ou c_j est *activée* et c est conforme à c_j), aucun triplet *incident* n'est généré.

L'état de la configuration c_j , dans la mémoire de l'automate, passera à l'état *validée* par une transition spécifique à partir d'un nœud contenant l'état *irradiation c_j* de la salle. Cette transition est associée au triplet (*accélérateur, irradiation, sans danger*), qui indique la fin de l'irradiation. Une opération réalisant le changement d'état est associée à la transition.

Une dernière opération importante à introduire est l'opération *vérifier(C)*. Cette opération vérifie si l'état des toutes les configurations est égal à *validée*. Lorsque c'est le cas, on génère un triplet *incident* qui, par une transition, permet d'arriver à un nœud contenant l'état *succès*.

Pour répondre aux besoins d'un système en radiothérapie, quelques contraintes supplémentaires sont posées sur la structure du modèle :

- un nœud contenant l'état *autorisation c_j* :
 - a une transition vers le nœud contenant l'état *irradiation c_j* ,
 - il existe au moins une transition vers ce nœud, à partir d'un autre nœud (en amont dans l'automate), effectuant l'activation de c_j ,
- un nœud contenant l'état *irradiation c_j* a au moins une transition, vers un autre nœud, effectuant la validation de c_j ,
- un nœud contenant l'état *conforme* :
 - ne doit pas avoir de transition vers un nœud contenant l'état *danger*,
 - il existe au moins une transition vers ce nœud, à partir d'un autre nœud (en amont dans l'automate), effectuant la vérification de la validation de toutes les configurations prédéfinies, ce qui implique qu'il y a au moins une transition vers un nœud contenant l'état *succès*,
- un nœud contenant l'état *attention* a au moins une transition vers un nœud contenant l'état *danger*,
- un nœud contenant l'état *danger* n'a aucune transition vers un autre nœud.

Comme une séance peut s'arrêter à tout moment, tout nœud doit permettre la sortie de l'automate, et chaque sortie représente un scénario. Les scénarios seront répartis en deux classes, selon qu'ils valident ou non la séance de traitement du point de vue du contrôle externe (par rapport aux critères présentés précédemment). Tout

nœud contenant un état *succès* représente un scénario qui valide la séance, et tout autre nœud ne la valide pas. À chaque visite d'un nœud, un signal visuel (interface graphique) est envoyé à l'utilisateur afin d'indiquer l'état courant de la salle.

Un exemple permettra de mieux comprendre le modèle présenté. Dans cet exemple, on considère que la salle est composée uniquement de deux éléments : l'accélérateur et le patient. L'accélérateur a uniquement un paramètre : *irradiation*, et le patient a deux paramètres : *présence* et *position*. Pour simplifier l'explication, les deux premiers paramètres ont des valeurs binaires (égales à *faux* à l'initialisation), et le dernier a des valeurs ternaires ($position = \{p_0, p_1, p_2\}$, et égal à p_0 à l'initialisation). Une seule configuration prédéfinie est donnée, et elle correspond à la présence du patient (*présence* égale à *vrai*) et à une position donnée (*position* égale à p_2). Faisons remarquer que par certains points cet exemple n'est pas réaliste car trop simpliste, mais qu'il permet de bien illustrer le potentiel du modèle sur des configurations plus complexes.

La Fig. 5.9 illustre l'automate définissant le modèle de scénario pour la validation d'une telle configuration. Nous avons utilisé des identifiants pour les triplets *incident* afin d'améliorer la lisibilité, ces identifiants étant composés de deux mots séparés du caractère « : ». Les différents événements possibles représentent les changements de valeurs des trois paramètres du système. Pour chaque nœud de l'automate, on doit comprendre :

- q_0 : aucune configuration n'a été validée et le patient est en dehors de la salle de traitement, il n'y a donc aucun risque,
- q_1 : le patient est l'intérieur de la salle de traitement et il risque d'être irradié incorrectement,
- q_2 : l'irradiation est en marche mais le patient n'est pas présent, il risque d'être irradié incorrectement s'il entre,
- q_3 : la configuration c_0 est *activée*, et on peut commencer à irradier,
- q_4 : l'irradiation est en marche pour la configuration c_0 , il ne faut pas que le patient se déplace,
- q_5 : le patient est irradié incorrectement,
- q_6 : la séance de traitement s'est déroulée comme prévu.

Bien que notre première approche ait été testée et validée, nous sommes convaincus du besoin d'inclure une analyse temporelle des événements dans notre module. Les aspects théoriques pour la deuxième approche ont été complètement définis, et elle est prête à être implémentée.

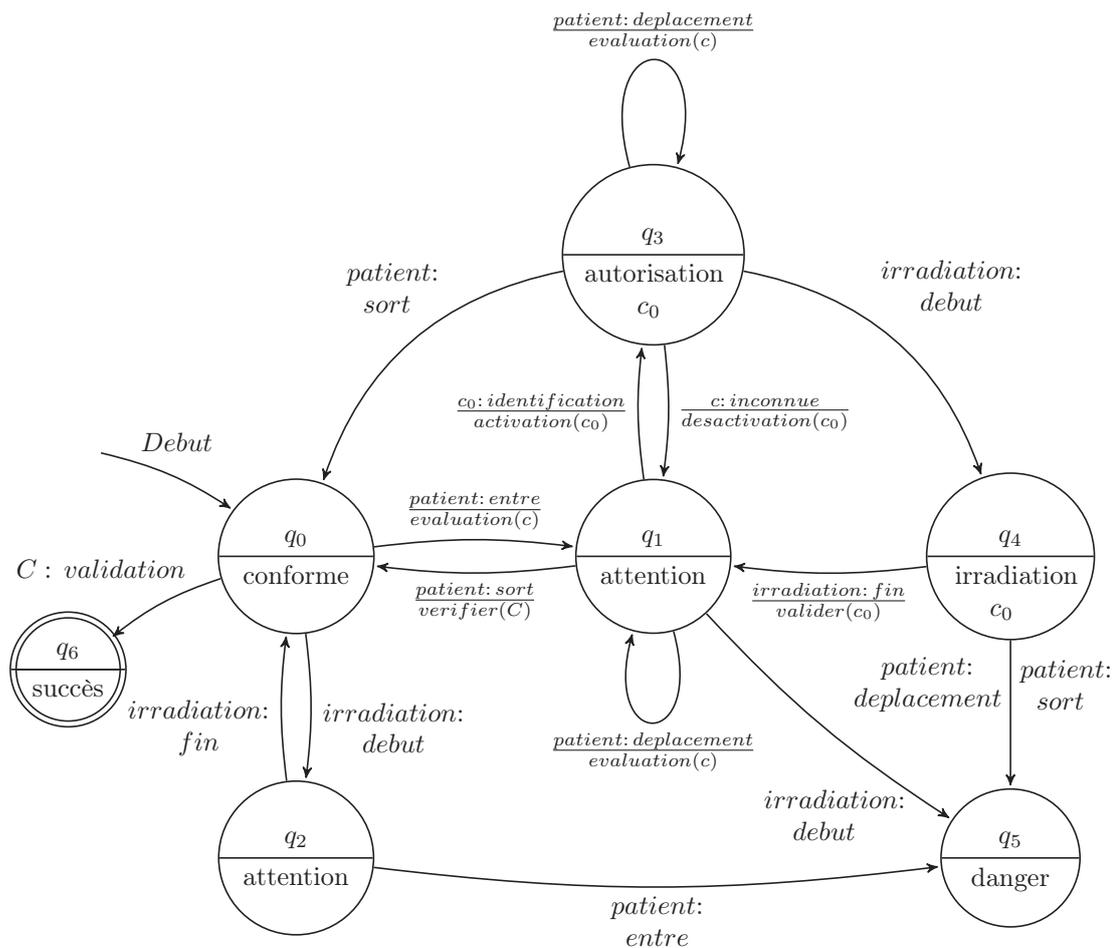


FIGURE 5.9: Exemple d'automate. $C = \{c_0\}$ est l'ensemble de configurations prédéfinies et c est la configuration courante.

Chapitre 6

Module de contrôle externe

Le module de contrôle est composé des cinq sous-modules présentés dans les chapitres précédents : initialisation du modèle, acquisition vidéo, suivi de mouvement, reconnaissance d'événements et reconnaissance de scénarios. Deux sous-modules supplémentaires permettent de le compléter : le premier, appelé *traçabilité*, est relatif à la sauvegarde des données extraites par les autres sous-modules ; le deuxième, appelé Interface Homme Machine, est relatif à l'interaction entre le module de contrôle et le personnel médical.

Ce chapitre clôture la présentation du module de contrôle externe pour la vidéosurveillance de salles de traitement en radiothérapie, avec la présentation de ces deux derniers sous-modules et une brève description du matériel et de la mise en place du module. Comme indiqué à la section 1.4, les sous-modules présentés dans ce chapitre ont été le résultat du travail d'Anthony PEREIRA, dans le cadre d'un stage de Master qui s'est déroulé sur les six derniers mois des travaux de cette thèse, sous l'impulsion et le financement de DOSIsoft.

6.1 Traçabilité

Il est indispensable, pour tout système d'assurance qualité, d'avoir un registre des informations acquises afin de garder une preuve du déroulement du processus. Dans notre cas, il s'agit notamment de sauvegarder :

1. les images acquises et les images virtuelles générées par ordinateur (donc, une paire d'images pour chaque caméra). Ces images sont enregistrées dans un flux vidéo, afin de profiter du codage vidéo pour assurer un meilleur taux de compression des données et ainsi réduire la taille occupée sur disque par la sauvegarde. Actuellement, l'enregistrement est fait au format MPEG-4 et, à titre d'exemple, un fichier vidéo de 300 mégaoctets (Mo) correspond à deux minutes d'enregistrement environ (donc, 600 Mo par caméra). La durée d'une séance de traitement, de la mise en place à l'irradiation, est d'une variabilité très large selon le type de cancer et la technique utilisée (de 15 min à plus d'une heure). En faisant un compromis sur la qualité des images (sur le principe qu'on ne fera pas d'analyse supplémentaire sur ces images) et en enregistrant les maillages des modèles numériques de la scène, au lieu des images virtuelles générées, on peut réduire considérablement la taille de la sauvegarde,

2. les paramètres de chaque élément à chaque instant t , qui sont enregistrés dans des « fichiers descripteurs ». Pour rappel, ces fichiers, au format XML, permettent d'enregistrer la description de la salle à un instant t . Afin d'éviter la redondance, de réduire la taille des fichiers et d'assurer (au mieux) la pérennité des fichiers, un dictionnaire est utilisé. Ce dictionnaire contient des identifiants courts (une lettre et deux chiffres dans la version actuelle) pour chaque élément et chaque type de paramètre. Actuellement, en enregistrant un fichier au format XML pour chaque instant t , l'ensemble des fichiers pour une durée de 2 minutes a une taille de 90 Mo. En utilisant une méthode incrémentale sur la différence des données des fichiers, cette sauvegarde peut être fortement réduite à moins de 5 Mo,
3. les événements détectés et reconnus à chaque instant t , qui sont enregistrés dans un fichier unique au format XML avec des balises bien définies. Le dictionnaire utilisé pour la sauvegarde des paramètres des éléments est également utilisé pour les événements; ce dictionnaire contenant également des identifiants pour les événements. Actuellement, pour deux minutes d'enregistrement, le fichier a une taille de moins de 5 Mo.

Pour un enregistrement de deux minutes, utilisant deux caméras, on utilise actuellement 1300 Mo environ d'espace disque. Dans le cas de l'utilisation de la reconnaissance de scénarios, cette taille reste environ la même. En effet, dans le pire des cas, chaque événement permet de reconnaître un scénario, et leur enregistrement sera fait dans le même fichier que les événements. Donc, c'est la taille du fichier d'événements qui sera doublée.

Le module de traçabilité s'exécute automatiquement en tâche de fond, parallèle aux autres modules, afin de réduire la durée des calculs à chaque instant t et d'assurer la sauvegarde de toutes les informations traitées et obtenues. Plus précisément, les informations acquises à l'instant $t - 1$ sont enregistrées pendant le traitement des informations à l'instant t (pour rappel, t est synchronisé sur la fréquence d'acquisition des images à partir du système de caméras).

6.2 Interaction avec l'utilisateur

Le personnel médical communique avec le module au travers d'une Interface Homme Machine (IHM) standard, c'est-à-dire, un écran, un clavier et une souris.

Du point de vue visuel, il s'agit d'une fenêtre de la taille de l'écran et contenant les informations sur les éléments et l'état global de la salle. La Fig. 6.1 présente une capture d'écran de cette fenêtre. Comme on peut le voir, les différentes fonctionnalités ont été regroupées en blocs :

l'outil de visualisation de l'état général de la séance est le témoin indiquant la conformité globale de la scène,

les écrans de vidéosurveillance permettent de visualiser les images acquises par les caméras et les images virtuelles générées (jusqu'à quatre caméras),

la liste des configurations prédéfinies, avec des informations sur la conformité des actions déjà réalisées par rapport aux faisceaux planifiés,

Outil de visualisation de l'état général de la séance

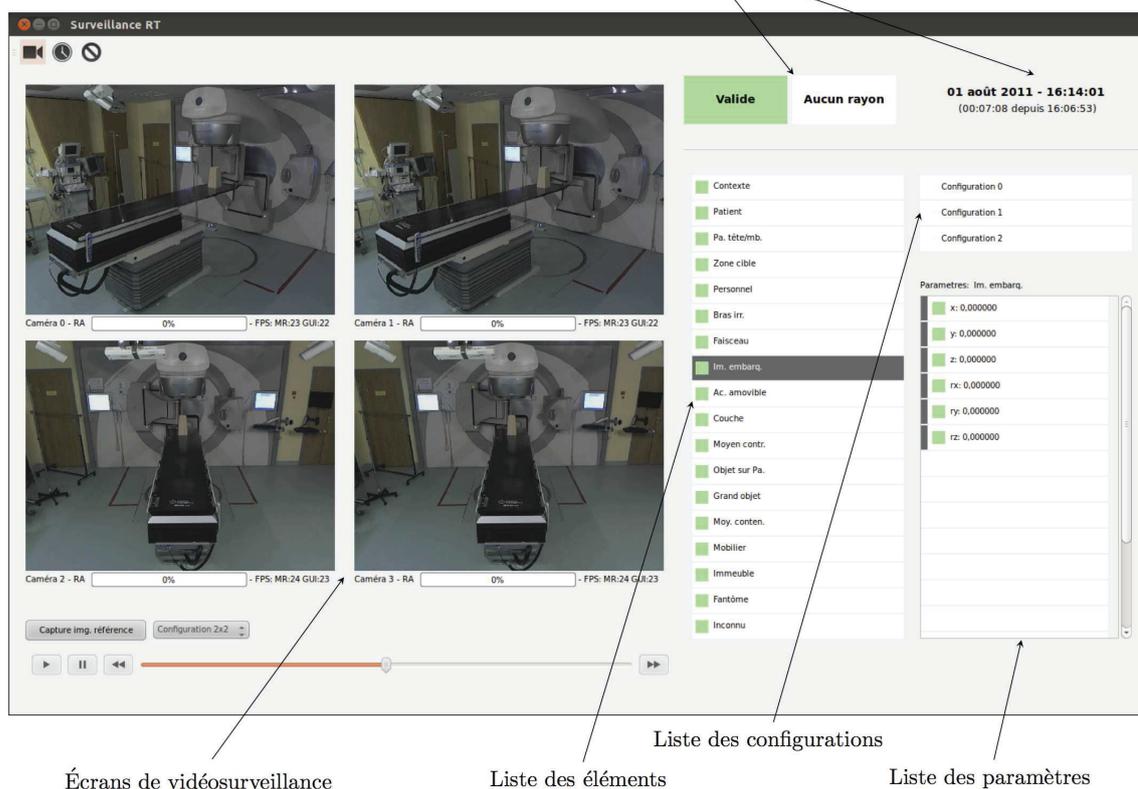


FIGURE 6.1: Capture d'écran de l'Interface Homme Machine du module de vidéosurveillance.

la **liste des éléments** montre l'état de chaque élément, ainsi que des informations plus précises sur sa configuration dans la **liste des paramètres**.

Nous avons établi un code couleur afin de bien différencier les niveaux de dangers. Ce code couleur est utilisé pour l'état global de la séance et pour l'état de chaque élément et de chaque paramètre de l'élément :

- : sans danger,
- : danger potentiel (zone de transition),
- : danger (arrêt immédiat de la séance).

Le choix d'utiliser uniquement un code couleur ternaire pour l'état de la scène, et de ne pas fournir des informations plus précises sur le taux de certitude entre la présence de danger ou pas (par exemple, une valeur allant de 0% à 100%), est justifié par l'importance de ces informations et la notion d'assistance au manipulateur. En effet, le manipulateur peut être perturbé dans l'analyse de la valeur donnée, se focalisant plus sur les chiffres que sur la véritable notion de danger. Nous avons donc décidé de réaliser cette analyse comme une tâche cachée à l'utilisateur, et se basant sur l'avis des experts (notamment par l'estimation des seuils).

6.2.1 État global de la séance

Le bloc contenant les informations sur l'état global de la séance montre les informations les plus importantes à connaître immédiatement lorsqu'on accède à l'écran de l'interface (Fig. 6.2, de gauche à droite) :

- l'état de la scène,
- l'état de l'irradiation,
- les informations temporelles : la date et l'heure du début de la séance ainsi que le temps écoulé.



FIGURE 6.2: Visualisation de l'état de la séance : deux exemples de configurations possibles.

Ces informations sont à jour à tout instant, et actualisées quand nécessaire. Dans le cadre indiquant l'état de la scène on utilise le code couleur (ternaire) présenté précédemment pour informer l'observateur sur la conformité du déroulement de la séance, alors que dans le cadre présentant l'état de l'irradiation on utilise un code de couleur binaire :

- : irradiation (en marche pour l'un des faisceaux),
- : non irradiation (aucun faisceau n'est administré).

6.2.2 Écrans de vidéosurveillance

Un écran de vidéosurveillance affiche les images réelles et virtuelles d'une caméra donnée (Fig. 6.3) et il dispose de deux outils :

1. le premier permet d'avoir une vue en réalité augmentée (Fig. 6.3(b)) par l'interpolation des images réelles et virtuelles. Il est possible de modifier le taux d'interpolation, et ainsi de se déplacer de la vue réelle jusqu'à la vue virtuelle, en passant par la vue en réalité augmentée,



(a) Sans réalité augmentée.

(b) Avec réalité augmentée.

FIGURE 6.3: Écrans de vidéosurveillance.

2. le deuxième outil exploite le modèle temporel du flux vidéo (cf. section 4.3.2) pour fournir une information visuelle sur la différence entre l'image courante

et une image dite « image de référence » (Fig. 6.4). Pour cela, il est possible d'enregistrer une image acquise pour chaque caméra, à un instant donné et à la demande de l'utilisateur. Ensuite, un passage en « mode référence » permet d'accéder aux informations sur cette différence, notamment en visualisant les pixels colorés en rouge qui correspondent à ceux considérés comme différents. Le principal intérêt de cet outil est de permettre la vérification visuelle que le patient ne s'est pas déplacé après la fin du positionnement, en particulier au niveau de la zone à irradier.

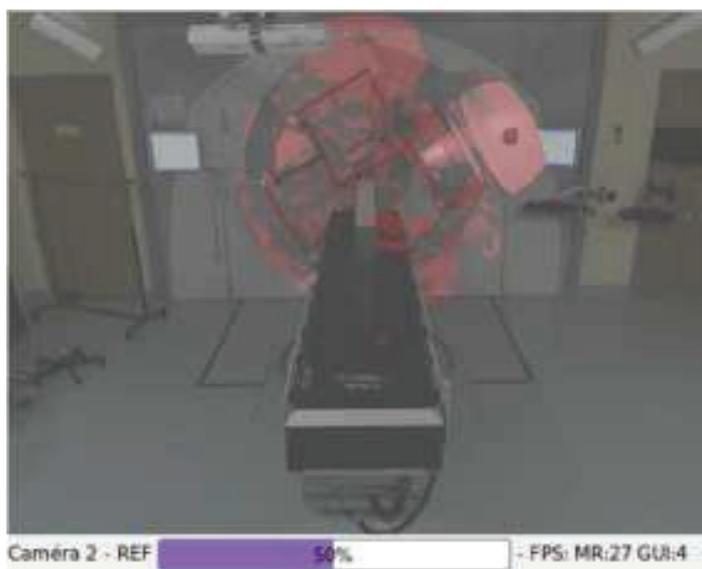


FIGURE 6.4: Écran de vidéosurveillance en mode référence.

6.2.3 Configurations

Dans la Fig. 6.5 nous avons regroupé les trois blocs contenant chacun une liste différente : liste des configurations, liste des éléments et liste des paramètres des éléments. La liste de configurations montre les configurations prédéfinies pour les faisceaux et donne des informations sur l'état de chacune suivant un code couleur :

- : l'irradiation a déjà eu lieu dans cette configuration, et aucune erreur n'a été détectée pendant cette irradiation,
- : la configuration a été détectée comme étant la plus proche (cf. section 5.3.2) de la configuration courante, mais l'irradiation n'a pas encore eu lieu dans cette configuration parce que la configuration courante n'est pas conforme,
- : la configuration n'est pas la plus proche de la configuration courante, et on n'a pas encore irradié dans cette configuration.

La liste des éléments montre ceux potentiellement présents dans la salle pour la séance de traitement en cours. Quatre informations sont données dans la ligne de chaque élément :

- l'état, à travers un cadre de couleur suivant le code couleur donné précédemment pour les niveaux de danger,
- l'identifiant associé à l'élément,

- l'heure du dernier événement détecté pour cet élément, s'il existe,
- l'identifiant du dernier événement reconnu pour cet élément, s'il existe.

Lorsqu'on clique sur un élément de la liste il dévient « actif ». Un seul élément peut être actif, et il est différencié visuellement des autres parce que sa ligne dans la liste a un fond de couleur gris foncé. La liste des paramètres est actualisée avec les informations sur les paramètres de l'élément actif. Pour chaque paramètre, quatre informations sont disponibles :

- l'état, à travers un cadre de couleur comme pour les éléments,
- l'identifiant associé au paramètre,
- la valeur courante du paramètre,
- la différence entre la valeur courante du paramètre et la valeur de référence contenue dans l'une des configurations des faisceaux non encore validées (la plus proche). Si la valeur courante est conforme, aucune valeur n'est affichée.

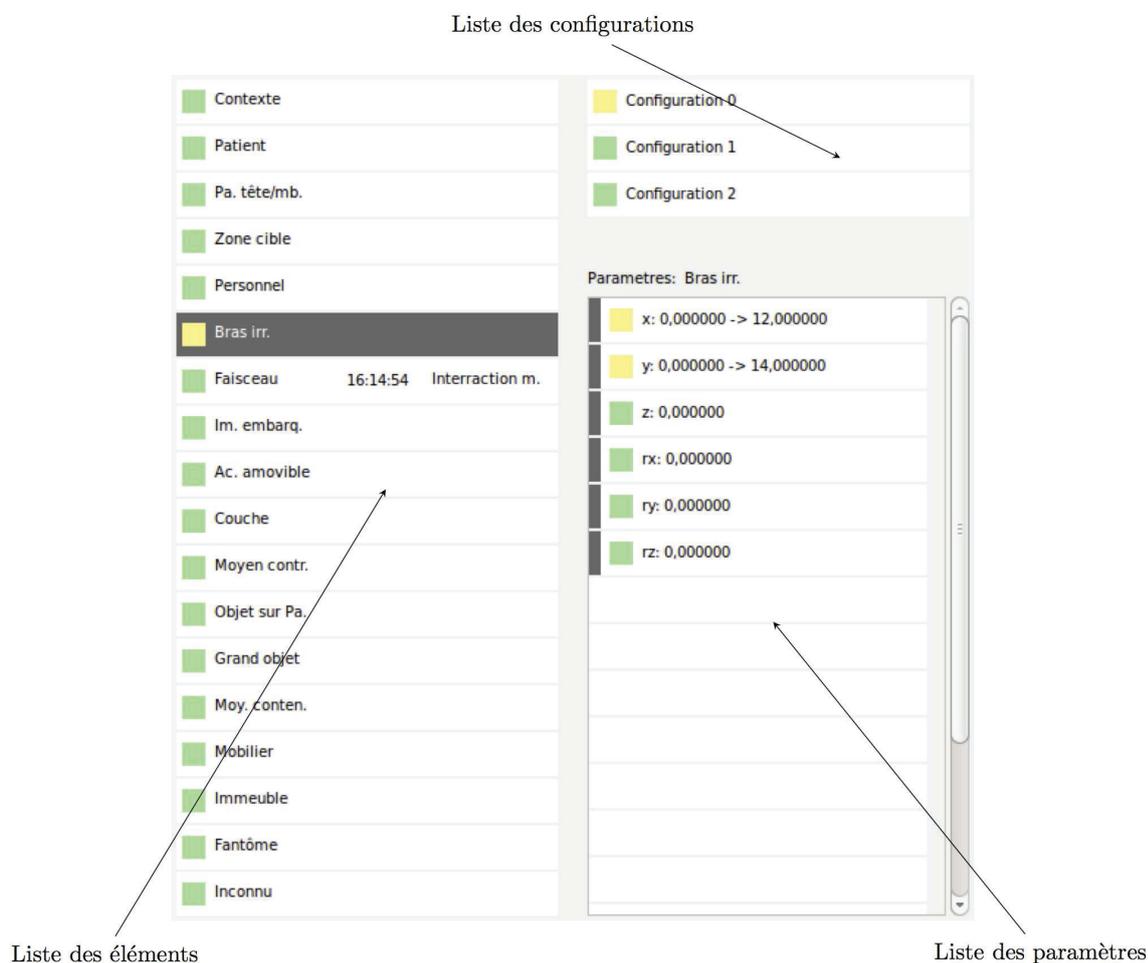


FIGURE 6.5: Listes des configurations, des éléments et des paramètres de ces éléments.

Dans le cas spécifique illustré dans la Fig. 6.5, on peut remarquer que les faisceaux ayant les configurations 1 et 2 ont été correctement délivrés (par rapport aux paramètres visuels externes). Pour que la configuration courante soit conforme à la configuration 0 (la seule qui manque), on doit modifier certains paramètres du bras d'irradiation, comme l'indiquent les carrés jaunes affichés après un clic sur la ligne de l'élément dans la liste. Les paramètres du bras d'irradiation sont visibles

car c'est l'élément actif, et dans la liste des paramètres on peut remarquer que les paramètres à modifier pour cet élément sont le paramètre x et le paramètre y (cadre jaune également).

Il est prévu à l'avenir que l'utilisateur puisse cliquer sur les images (écrans de vidéosurveillance) afin d'accéder directement aux informations relatives à un élément. Cela sera possible grâce aux masques de visibilité des éléments.

6.2.4 Frise chronologique

Une deuxième fenêtre permet à l'utilisateur de visualiser l'évolution dans le temps de la scène (Fig. 6.6). On peut voir, pour chaque élément, la succession d'états par lesquels il est passé au cours de la séance. Cela est représenté par une ligne de couleur (de gauche à droite) respectant le code couleur des niveaux de danger présenté précédemment. En cliquant sur un point de cette ligne, une « info bulle » contenant l'identifiant de l'événement responsable de l'état de l'élément au point où l'on a cliqué apparaît.

L'échelle de la frise chronologique peut être modifiée. Il est donc possible de se focaliser sur une zone pour observer l'évolution détaillée dans un intervalle de valeurs de t , ou bien il est possible de voir le déroulement de la séance d'un point de vue plus global (par exemple, voir l'évolution des éléments sur toute la durée de la séance avec moins de détails – cas de la Fig. 6.6).



FIGURE 6.6: Frise chronologique.

6.3 Interaction des sous-modules

Les sous-modules interagissent à travers le modèle numérique, contenant toutes les informations de la scène, comme présenté sur la Fig. 6.7. Au niveau des images et

des paramètres des éléments, on a uniquement besoin des informations à l'instant t pour la vidéosurveillance de la scène. Les informations à l'instant $t - 1$ sont gardées temporairement afin de pouvoir être enregistrées parallèlement au traitement des informations à l'instant t . Au niveau des événements et des scénarios, les informations sur toute la durée doivent être disponibles à tout instant. Cela ne pose pas de problème particulier, car on utilise peu d'espace dans la mémoire pour ces informations (seulement quelques mégaoctets). Les images occupent également quelques mégaoctets, mais les modèles numériques des éléments (paramètres, modèles géométriques, etc.) occupent une place importante en la mémoire (de l'ordre de 200 Mo pour deux caméras et trois éléments).

L'ordre d'exécution, comme présenté à la Fig. 3.1 (cf. section 3.1), et la synchronisation des sous-modules assurent la disponibilité des informations nécessaires à l'exécution des tâches de chacun des sous-modules.

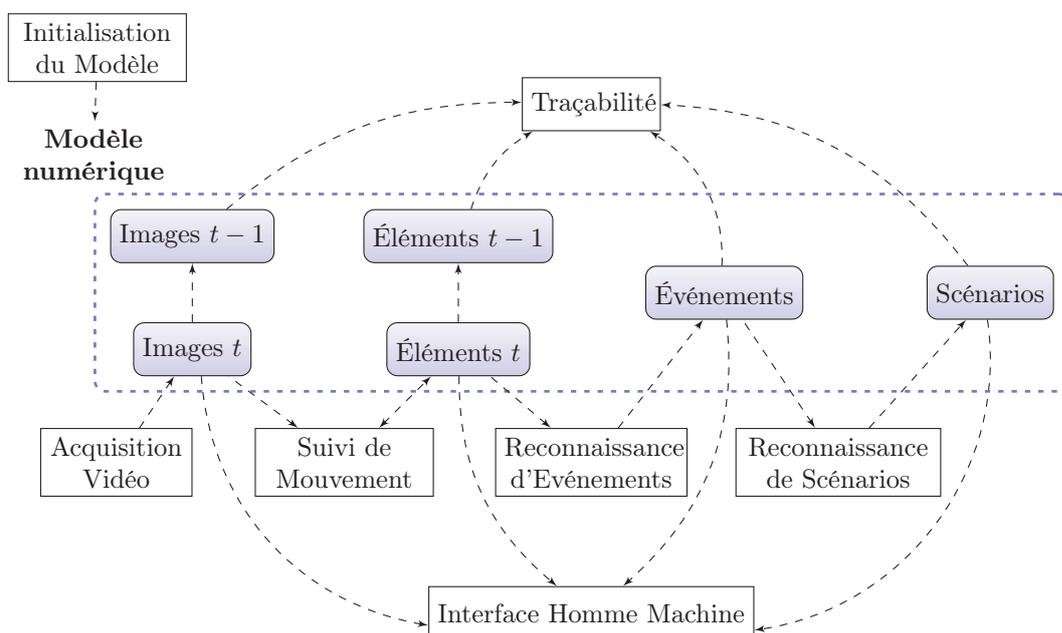


FIGURE 6.7: Interaction des sous-modules.

6.4 Implémentation du module

La partie logicielle du module de vidéosurveillance a été développée dans un environnement Unix et en langage C++, ce choix étant justifié pour la disponibilité des bibliothèques logicielles suivantes :

OpenCV (Open Source Computer Vision) avec des fonctions pour le traitement d'images en temps réel [22],

Coin3D avec des fonctions pour la gestion des graphes de scènes en trois dimensions [23]. Cette bibliothèque fournit une surcouche de programmation pour OpenGL (Open Graphics Library) [24], cette dernière contenant des fonctions pour la génération d'images à partir d'une description numérique d'un monde virtuel,

Xerces-C++ avec des fonctions pour le traitement (lecture, écriture, validation, modification, etc.) de fichiers au format XML [25],

DCMTK (DICOM ToolKit) avec des fonctions pour la traitement de fichiers au format DICOM et d'autres fonctions pour l'échange des données dans ce format [26],

QT avec des fonctions pour la création des Interfaces Homme-Machine [27].

Pour la partie matérielle, il s'agit d'un équipement standard (ordinateur avec écran, clavier et souris) et des webcams (système d'acquisition).

6.5 Mise en place du module

Afin de pouvoir suivre de près ce qui se passe à l'intérieur de la salle de traitement lors des tests, nous avons utilisé un ordinateur portable, placé dans la salle, auquel nous connectons les caméras placées dans la même configuration que les caméras actuellement présentes (cf. section 5.1.1). Les tests ont donc été réalisés sans mettre en marche l'irradiation, et l'irradiation a été simulée pour les tests au niveau des événements. Cependant, le système a été conçu pour suivre la configuration globale déjà existant dans la salle : les caméras à l'intérieur de la salle sont connectées à un ordinateur fixe placé dans le poste de commande. Une démarche souvent utilisée consiste à enregistrer des images acquises des caméras et réaliser le traitement de ces séquences d'images *a posteriori*. Avec cette démarche, on est capable de réaliser des tests sur un ordinateur fixe.

Toute la partie logicielle du module est installée sur un seul ordinateur. Les fichiers nécessaires sont actuellement transférés à l'aide d'un support de type clef USB (moins de 4 gigaoctets), et les données sont exportées sur ce même type de support ou sur un disque dur externe lorsqu'il s'agit de fichiers de grande taille. Dans le futur, et dans l'idéal, le système doit être connecté au réseau du service de radiothérapie du centre hospitalier afin de faciliter l'importation et l'exportation des données.

Le calibrage intrinsèque et extrinsèque des caméras est réalisé grâce aux outils de la bibliothèque logicielle OpenCV. Comme expliqué à la section 3.2.1, une alternative au calibrage extrinsèque est d'utiliser un dispositif muni de repères visuels pour ce calibrage et permettant de connaître la relation entre le repère de la salle à l'isocentre et le repère de calibrage. Ce dispositif facilite donc le calibrage extrinsèque dans le repère de la salle.

Dans le chapitre 4, nous avons introduit le recalage 3D et nous avons expliqué comment calculer la pose 3D d'un élément. Dans cette technique, il s'agit donc de calculer la relation entre le repère de la caméra et le repère de l'élément. Dans le cas du suivi du mouvement des éléments, le repère de l'élément change à chaque déplacement et on calcule la nouvelle relation entre les deux repères. Le contraire peut être également réalisé, c'est-à-dire, on peut déplacer la caméra « virtuellement » et calculer la transformation par rapport au repère d'un ou plusieurs éléments qui ont resté immobiles. Bien qu'il ne s'agisse pas de déplacer réellement la caméra, ce principe peut être utilisé pour calculer la pose 3D de la caméra par rapport au repère d'un élément, principe déjà proposé et utilisé dans le domaine de la réalité



FIGURE 6.8: Dispositif de calibration extrinsèque des caméras tenant compte du repère à l'isocentre.

augmentée, le meilleur exemple étant dans l'exploitation de la bibliothèque logicielle ARToolKit [20], qui est utilisée dans l'exemple de la Fig. 3.3(b).

Afin de rester homogène, on utilise la méthode présentée au chapitre 4 pour réaliser le calibration extrinsèque des caméras. Pour cela, nous utilisons une boîte en forme de pyramide tronquée (Fig. 6.8). Un premier positionnement manuel des caméras, à l'aide d'un écran affichant une vue en réalité augmentée, permet de respecter les prérequis de la méthode. Ensuite, la méthode de recalage 3D implémentée est utilisée pour obtenir les paramètres extrinsèques de caméras.

6.5.1 Salle de traitement et modèles numériques des éléments

Comme indiqué tout au long du manuscrit, nous avons utilisé une salle située au Centre Léon Bérard [7], à Lyon, pour réaliser nos tests. Plusieurs illustrations de cette salle ont déjà été présentées : Fig. 1.4 (page 7), Fig. 3.7 (page 51), Fig. 4.2 (page 58), Fig. 4.3 (page 59), Fig. 5.1 (page 82) et Fig. 6.8.

Pour les tests du module, nous nous sommes focalisés sur cinq éléments et nous avons réalisé des modèles géométriques de ces éléments à l'aide du logiciel Blender¹ :

1. La boîte « pyramidale », utilisée comme dispositif de calibration extrinsèque (Fig. 6.8 et Fig. 6.9), sert aussi de « fantôme » d'élément quelconque afin de réaliser des tests sur les interactions entre les éléments (par exemple, un élément sur la table qui se déplace à cause du déplacement de la table) ou sur des situations au niveau de la zone à irradier du patient (par exemple, mauvais positionnement).
2. Un objet de forme simple (parallélépipède rectangle) et fait d'un matériel léger (polystyrène expansé), utilisé pour tester les collisions entre deux éléments.
3. L'accélérateur de particules (Fig. 2.7(b) et Fig. 6.10), qui correspond au modèle Synergy® System du constructeur Elekta [14]. Il a notamment deux degrés de liberté : une rotation de tout le dispositif autour de l'axe médian du disque (au

1. Blender est un logiciel d'animation, de modélisation et de rendu 3D [28], qui permet l'importation et l'exportation de fichiers au format VRML, ce format étant également pris en charge par les outils de rendu de la bibliothèque logicielle Coin3D [23] que nous avons utilisée.

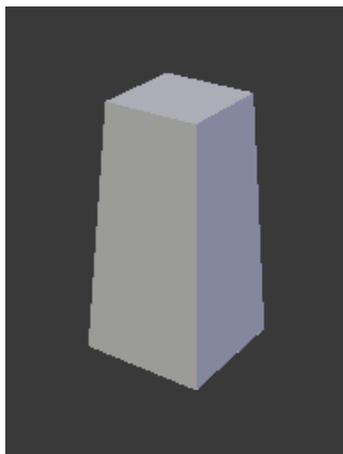
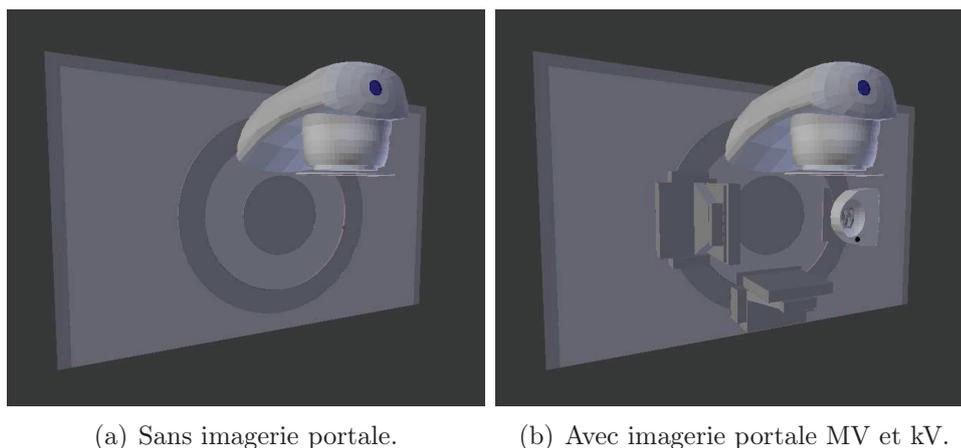


FIGURE 6.9: Modèle géométrique numérique du dispositif de calibration externe des caméras.

niveau du mur) supportant le bras d'irradiation, et une rotation d'une partie de l'accélérateur (partie grise métallique au niveau de la « tête de l'accélérateur ») autour de l'axe du faisceau d'irradiation. Nous avons également modélisé les éléments d'imagerie portale (cf. section 2.1.1) couplés à l'accélérateur, que nous comptons ajouter dans des tests futurs.



(a) Sans imagerie portale.

(b) Avec imagerie portale MV et kV.

FIGURE 6.10: Modèle géométrique numérique de l'accélérateur de particules.

4. La table de traitement (Fig. 2.11(c) et Fig. 6.11), qui correspond au modèle de table HexaPODTMevo RT System du constructeur Elekta [14]. Cette table a six degrés de libertés : une rotation au niveau de la base sur le sol et une translation en hauteur de tout le dispositif (axe z), et deux rotations et deux translations de la partie haute de la table sur les autres axes (x et y).
5. Le patient (Fig. 6.12). Nous avons créé un modèle géométrique numérique sur la base qu'il puisse être utilisé pour tout individu potentiellement présent dans la salle, cela à l'aide d'un outil permettant d'adapter le modèle géométrique à chaque cas spécifique. Ce modèle, en « position zéro » (à droite de la Fig. 6.12), est composé de coupes parallèles en hauteur. Ce choix a été fait afin de faciliter la comparaison entre les informations de ce modèle et les informations issues de l'étude d'imagerie au scanner sur le patient : coupes réalisées dans le même axe et le patient est généralement placé dans la même configuration.

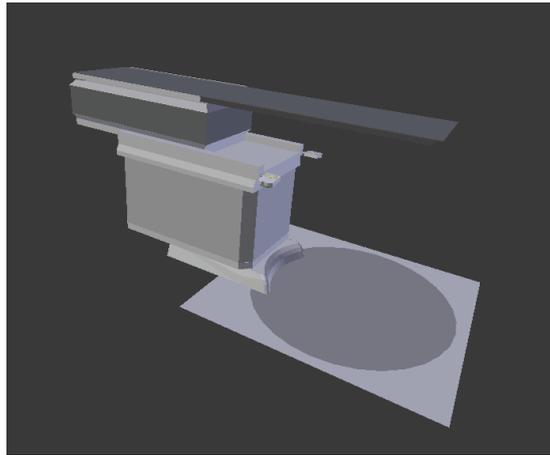


FIGURE 6.11: Modèle géométrique numérique de la table de traitement.

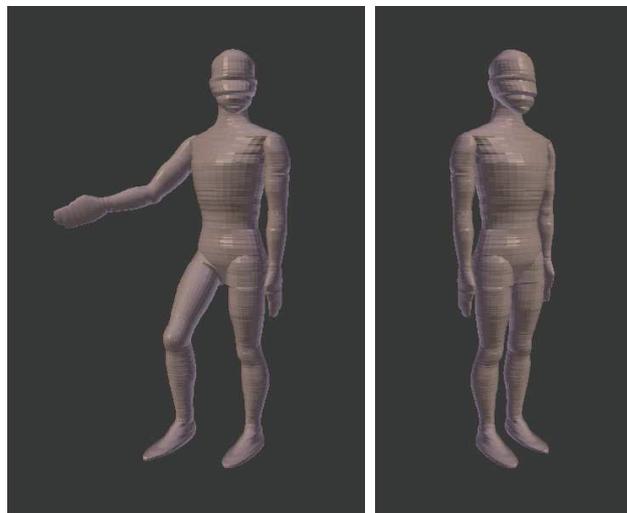


FIGURE 6.12: Modèle géométrique numérique générique d'un individu.

Pour ces éléments nous avons construit des fichiers descripteurs contenant des informations sur ces éléments : degrés de liberté, nom de fichier contenant le modèle géométrique (au format VRML), les informations sur les descripteurs, etc.. Enfin, un fichier regroupe toute la description de la salle (en incluant des liens sur les fichiers de chaque élément), et les informations sur la salle à vide (murs, sol, plafond, etc.). L'initialisation de la configuration de ces éléments et de leurs modèles numériques (incluant les modèles géométriques) est réalisée comme présenté à la section 3.3.6, et l'environnement de réalité augmentée est obtenu comme présenté à la section 3.4.

Chapitre 7

Validation et résultats

Ce chapitre est consacré à la présentation des résultats des trois sous-modules qui ont été décrits tout au long de ce document : initialisation du modèle, suivi de mouvement et reconnaissance d'événements. Les résultats sont analysés, à l'aide de graphes, de figures et/ou de tableaux dans certains cas, et des conclusions sur ces résultats sont données.

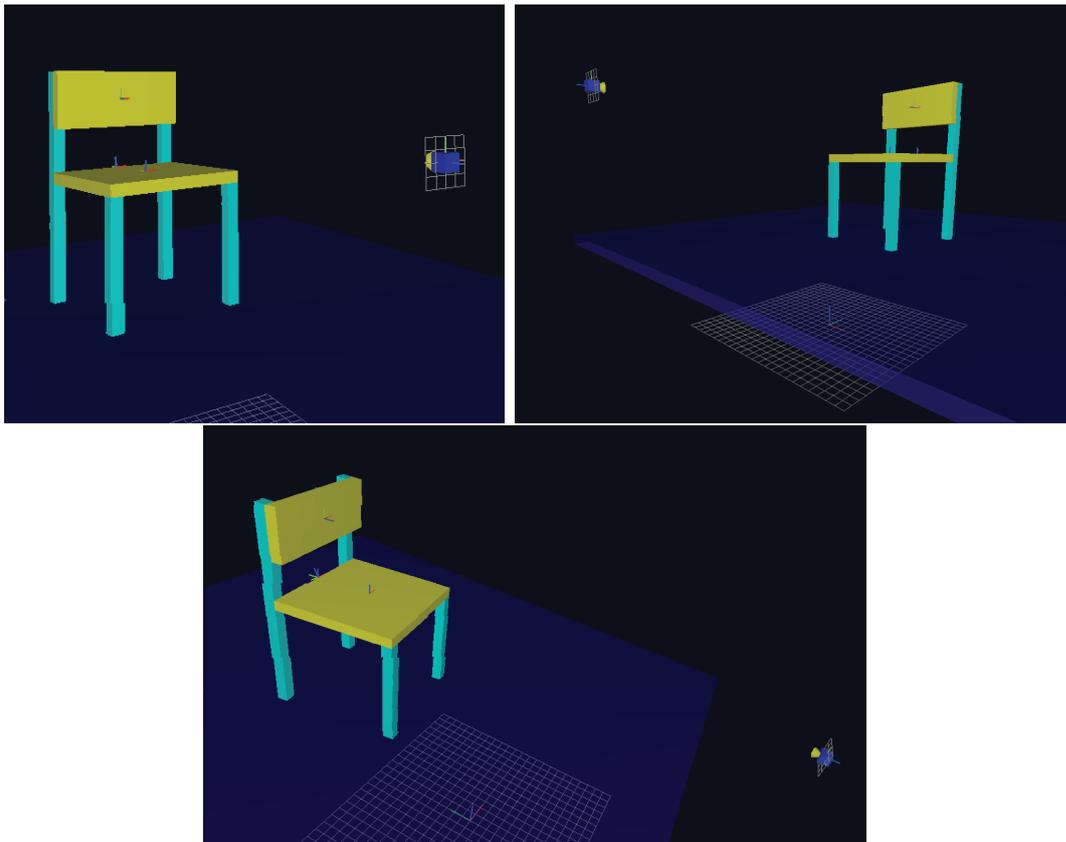
7.1 Modèle générique

Une première modélisation (réalisée dans le cadre du stage de Master précédant cette thèse – cf. section 1.4) d'une scène simple contenant trois éléments : le sol, une chaise et une caméra, nous a permis de vérifier les capacités du modèle générique présenté à la section 3.3. Pour cette modélisation, nous avons utilisé des marqueurs de la bibliothèque logicielle ARToolKit [20] (cf. section 2.4.2) pour le suivi de mouvement afin de pouvoir tester le modèle générique de manière indépendante les sous-modules du système de vidéo-surveillance. Des fonctions contenues dans cette bibliothèque permettent de calculer la pose 3D de chaque marqueur par rapport à une caméra.

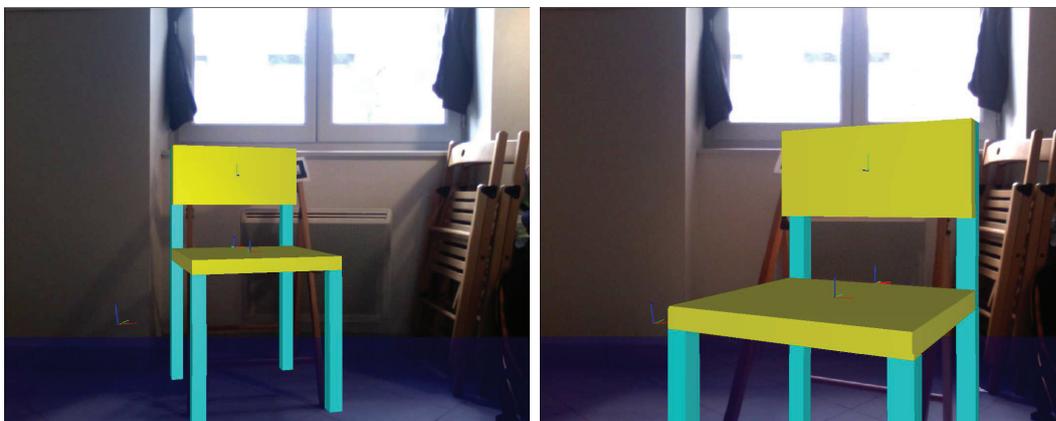
Les marqueurs ont été utilisés uniquement pour la chaise, et pour éviter les problèmes liés à l'occultation des motifs, nous avons collé ces marqueurs dans la partie supérieure du dos de la chaise. Les marqueurs ont été définis comme des descripteurs de la chaise dans le modèle. Plusieurs contraintes ont été ajoutées à la scène, notamment au niveau des degrés de liberté de la chaise. Nous distinguons une contrainte spéciale, qui implique deux éléments : la chaise et le sol. Cette contrainte force tous les pieds de la chaise à être toujours en contact avec le sol, et elle est traduite dans le modèle par un opérateur fixant la distance entre ces deux éléments.

Cette scène simple nous a permis de démontrer la faisabilité du modèle générique et de jeter les bases d'un système de suivi de mouvement fondé sur des modèles numériques des éléments réels à suivre. C'est notamment grâce à la notion de descripteurs (cf. section 3.3.4) que les mouvements de la chaise ont pu être calculés, même lorsque l'un des deux motifs était caché ou partiellement visible.

L'utilisation et l'utilité des opérateurs ont été également validées par des tests. L'opérateur de « contact » a permis de lever des ambiguïtés, notamment celles dues au bruit d'acquisition. En effet, de petites variations dans l'image peuvent provoquer des résultats parfois peu réalistes (par exemple, les petites déplacements continus



(a) Environnement virtuel



(b) Environnement de réalité augmentée

FIGURE 7.1: Résultats visuels de la modélisation d'une scène simple.

qui donnent l'impression que la chaise tremble et/ou qu'elle s'enfonce dans le sol). La contrainte exprimée à travers cet opérateur a permis de réduire l'espace de solutions et de ne considérer que les solutions réalistes, améliorant les résultats. En général, les contraintes incluses dans le modèle permettent d'améliorer les résultats du suivi de mouvement des éléments par rapport aux résultats obtenus quand les contraintes ne sont pas exploitées. Enfin, cette première modélisation a démontré empiriquement l'utilité et les capacités d'utilisation de fichiers descripteurs pour la sauvegarde du modèle numérique. Des résultats visuels sont présentés à la Fig. 7.1.

7.1.1 Modélisation d'une salle de traitement

Le modèle générique nous a permis de créer le modèle numérique d'une salle de traitement, comme présenté à la section 6.5. Dans ce contexte, les contraintes liées au traitement et au facteur humain impose qu'on laisse de coté (mais pas complètement) l'utilisation de marqueurs. Pour cette modélisation, on s'est donc focalisé sur l'utilisation de la méthode de suivi de mouvement présentée dans le chapitre 4, qui exploite les motifs déjà présents sur les différents éléments (par exemple, des logos sur la surface des machines). Les descripteurs sont alors définis sur la base de ces motifs.

À cette occasion, nous avons testé avec succès les capacités des outils d'importation des données à partir d'autres fichiers :

- au format XML, contenant la description du modèle numérique d'un ou plusieurs élément (avec récursivité),
- au format DICOM, contenant principalement les informations sur le plan de traitement,
- au format VRML, contenant les modèles géométriques des éléments dans la salle.

Un fichier descripteur unique contient la description complète de la salle, et fait appel aux autres fichiers.

Enfin, c'est cette modélisation qui a été utilisée pour les tests présentés dans la suite de ce chapitre portant sur les sous-modules du système de vidéosurveillance. Des résultats visuels sont présentés à la Fig. 7.2.

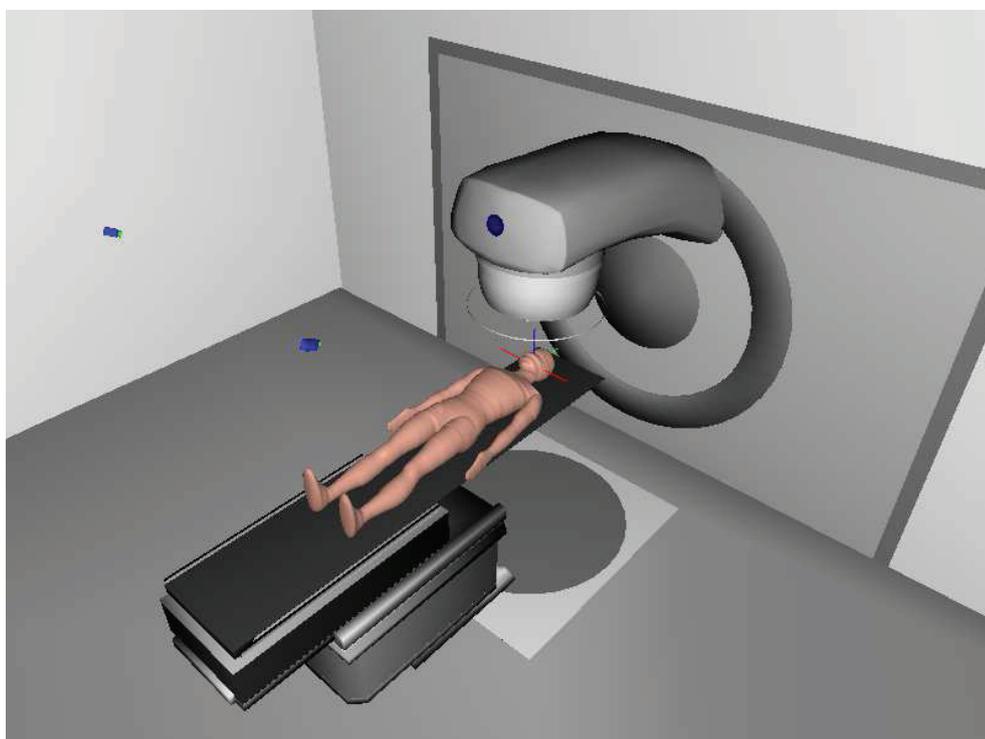


FIGURE 7.2: Résultats visuels de la modélisation d'une salle de traitement.

7.1.2 Conclusions sur le modèle générique

La particularité du modèle générique est qu'il doit être capable de fournir des informations de différents types aux différents sous-modules (cf. section 6.3). Il touche donc à des sujets de recherche très actifs, comme le suivi précis de mouvements basé objet et la modélisation fidèle d'humanoïdes.

Les tests réalisés ont démontré que le modèle présenté était suffisamment générique pour s'adapter aux différentes situations jusqu'à maintenant rencontrées, et plus particulièrement aux configurations d'une salle de traitement par radiothérapie.

L'utilisation de fichiers pour chaque élément est une première étape dans le transfert d'informations entre deux séances consécutives dans une salle de traitement (patients différents). À la fin d'une séance, la description numérique de chaque élément est sauvegardée, et au début de la séance suivante on réutilise les informations des éléments encore présents dans la salle (tous sauf le patient et les moyens de contention dans le cas général) afin de personnaliser la description de la séance pour le prochain patient.

La construction actuelle du modèle numérique de la salle est faite sur un ensemble générique d'éléments, c'est-à-dire les mêmes éléments sont considérés pour toute séance (avec une personnalisation pour cette séance). Nous prévoyons, dans une version plus avancée du système de vidéosurveillance, que le personnel médical soit capable de choisir, dans une bibliothèque prédéfinie, les éléments qui seront présents dans le modèle numérique de la séance.

Cette partie du système a fait l'objet d'une publication dans une conférence internationale : [PDM11].

7.2 Suivi de mouvement

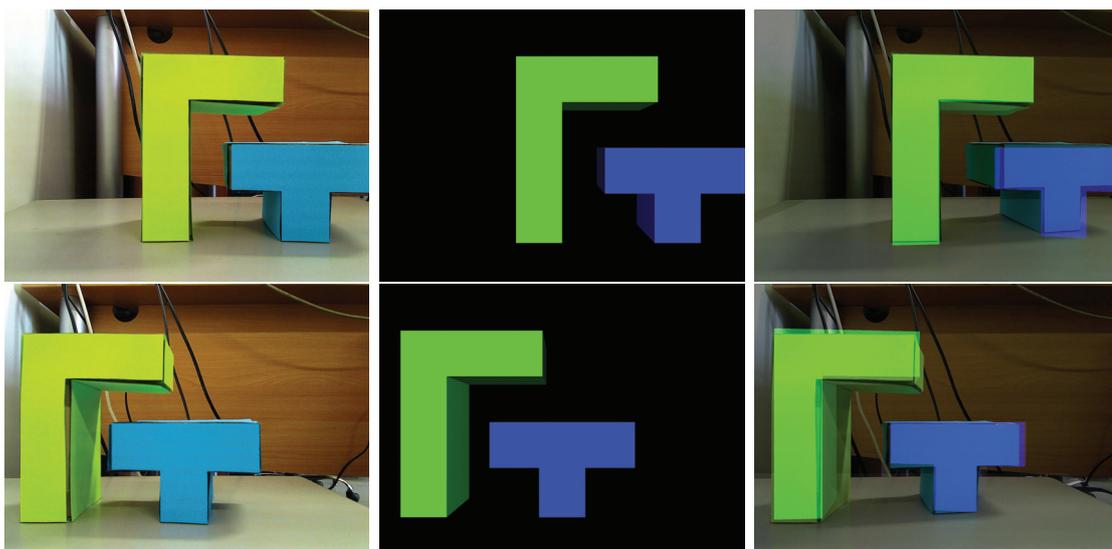
Les premiers tests de la méthode de suivi de mouvement présentée dans le chapitre 4 ont été réalisés en dehors des salles de traitement par radiothérapie, mais sur des scènes équivalentes. Ils ont notamment servi au développement et à l'évolution de la méthode et la mesure (par les résultats) des choix et des améliorations possibles. Quelques illustrations sont présentées à la Fig. 7.3.

Une fois la méthode arrivée à un état assez avancé de développement, nous avons commencé à réaliser des tests dans une vraie salle de traitement au Centre Léon Bérard [7]. Le scène a été décrit comme étant composée de la salle à vide (murs, sol, etc.), de deux éléments toujours présents : le bras d'irradiation et la table de traitement, et de quatre éléments potentiellement présents : deux individus (patient et personnel) et deux objets de forme simple représentant un objet quelconque (cf. section 6.5). À chaque test réalisé, tous les éléments n'ont pas été pris en compte, mais la description de la salle restait la même.

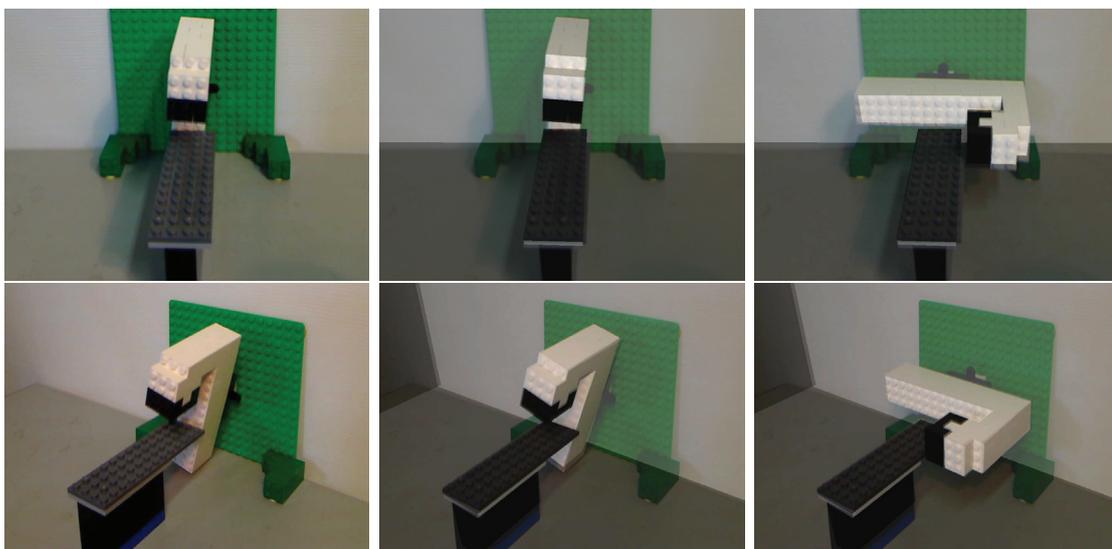
Toutes les évaluations présentées dans la suite de ce chapitre ont été réalisées avec deux caméras, mais le système pourrait fonctionner avec plus.

7.2.1 Résultats des tests sur des données synthétiques

Nous présentons d'abord des résultats de la méthode sur des données synthétiques. Nous avons généré des séquences d'images par ordinateur afin d'obtenir une



(a) Utilisation d'objets de forme très simple représentant l'accélérateur et la table de traitement. Une ligne par caméra. Gauche : image réelle à l'initialisation ($t = 0$) ; centre : image virtuelle du modèle numérique à l'initialisation ; droite : image en réalité augmentée à un instant $t \neq 0$.



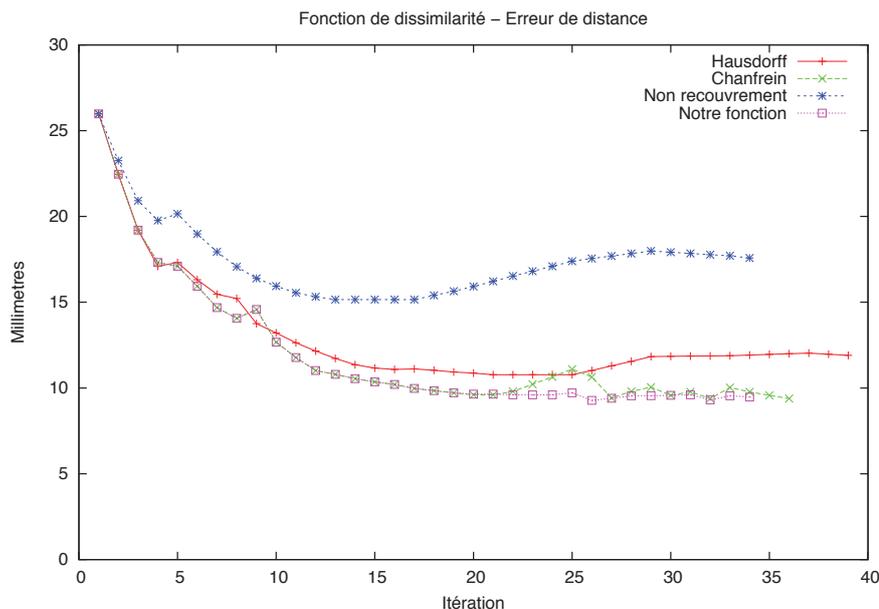
(b) Utilisation de legos® pour créer des formes plus complexes représentant l'accélérateur et la table de traitement. Une ligne par caméra. Gauche : image réelle à l'initialisation ($t = 0$) ; centre : image en réalité augmentée à l'initialisation ; droite : image en réalité augmentée à un instant $t \neq 0$.

FIGURE 7.3: Résultats visuels des premiers tests de la méthode de suivi de mouvement.

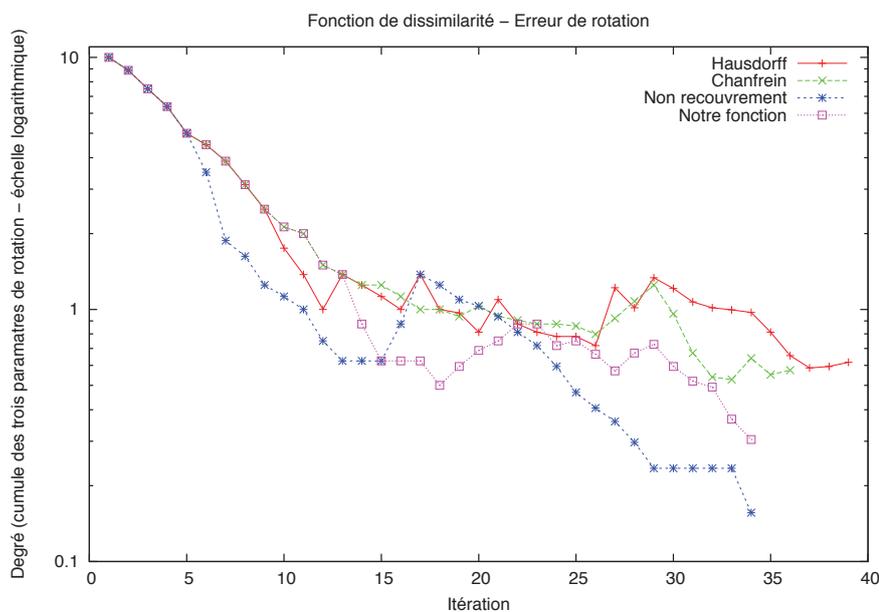
vérité terrain la plus précise possible et avec très peu du bruit (voire, pas du tout). Ces séquences d'images sont des simulations d'acquisition à partir de caméras, et elles sont obtenues grâce aux modèles numériques de caméras et de la scène (cf. section 3.2). Afin d'obtenir des évaluations plus génériques, certains éléments ont été dupliqués et placés dans des positions arbitraires. La seule contrainte établie pour les éléments de la scène est que chaque élément soit visible, au moins partiellement, par le système de caméras à tout instant. Les degrés de liberté ont également subi des changements indépendamment de leur description réelle. En général, tous les éléments présents dans les images synthétiques ont six degrés de libertés (trois

translations et trois rotations). Enfin, les images ont été générées avec une résolution en pixels de 640×480 .

Pendant la génération des images synthétiques, chaque élément est déplacé dans une pose 3D choisie aléatoirement à chaque instant t . Ces déplacements simulés ont une amplitude maximale de 2 cm et 5 degrés pour les différents paramètres de la pose, afin d'assurer une « dynamique homogène » de la scène et créer des conditions proches des caractéristiques de la scène réelle. À chaque instant t , les poses 3D de tous les éléments sont enregistrées et ensuite une image est générée.



(a) Erreur de distance entre la pose simulée et la pose calculée.



(b) Erreur de rotation entre la pose simulée et la pose calculée.

FIGURE 7.4: Graphes pour l'évaluation de notre fonction de dissimilarité lors de la procédure d'optimisation pour le calcul de la pose à un instant t .

La méthode de suivi de mouvement a été exécutée sur ces séquences d'images, et les résultats des poses 3D calculées ont été comparés aux données enregistrées lors de la génération des images. Finalement, nous avons évalué les différences entre les données simulées et les données calculées.

Les graphes de la Fig. 7.4 présentent des résultats permettant de comparer les différentes fonctions de dissimilarité introduites à la section 4.4. Dans ces graphes on peut voir l'évolution de l'erreur, en translation (Fig. 7.4(a)) et en rotation (Fig. 7.4(b)), entre la pose courante de l'élément (pose du modèle numérique) et la pose cible (pose de l'objet réel-simulé) lors du calcul de la pose à un instant t . On peut voir, plus particulièrement, comment l'erreur est réduite grâce à la méthode d'optimisation de chaque fonction de dissimilarité utilisée.

Ces résultats montrent que notre fonction de dissimilarité est globalement plus performante que les autres fonctions déjà publiées. Lorsqu'on considère l'aspect distance et rotation séparément, notre fonction n'est pas forcément la meilleure mais elle reste comparable aux autres méthodes. Cependant, notre fonction est celle qui donne les meilleurs résultats lorsqu'on fait un compromis entre les erreurs de distance et de rotation à la fois.

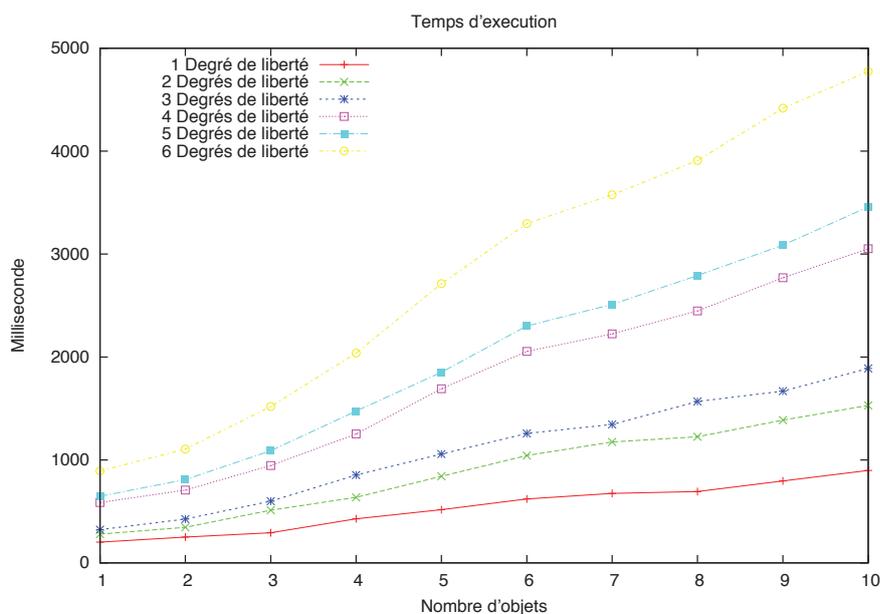


FIGURE 7.5: Adaptabilité aux nombre d'éléments par rapport au temps de calcul à un instant t de la méthode de calcul de la pose 3D.

Le graphe de la Fig. 7.5 présente des résultats permettant d'illustrer l'adaptabilité au nombre d'éléments de la méthode par rapport au nombre d'objets traités. Dans ce graphe on peut voir le temps de calcul pour tous les éléments à un instant t en fonction du nombre d'éléments et du nombre de degrés de liberté (même nombre de degrés pour tous les éléments considérés pour chaque évaluation). On peut voir que la méthode a un comportement linéaire par rapport au nombre d'éléments. La non linéarité qu'on peut observer lorsque le nombre d'éléments est petit est due au fait que certains calculs ne sont pas réutilisés, en particulier la génération d'images globales de la scène. Ces résultats sont encourageants car il est encore possible de les améliorer par la parallélisation de la méthode (dans des étapes de la méthode,

mais aussi dans le calcul parallèle de la pose de plusieurs éléments) et l'utilisation de la programmation graphique (puce graphique dédiée).

Le graphe de la Fig. 7.6 présente un exemple d'exécution dans le temps de la méthode de suivi de mouvement par calcul de la pose 3D. Dans ce graphe il est possible de voir l'évolution de l'erreur de distance (translation) entre la pose simulée et la pose calculée. Comme on peut le remarquer, l'erreur de calcul ne s'accumule pas et la méthode est capable de calculer une bonne approximation de la pose à un instant t , même quand le calcul à l'instant précédent $t - 1$ a une erreur importante (par rapport à la moyenne et à l'écart type de l'erreur dans le temps). Le comportement local pourrait être amélioré par la mise en place d'une stratégie basée sur les poids des poses calculées aux instants précédents, sur une fenêtre de longueur fixe.

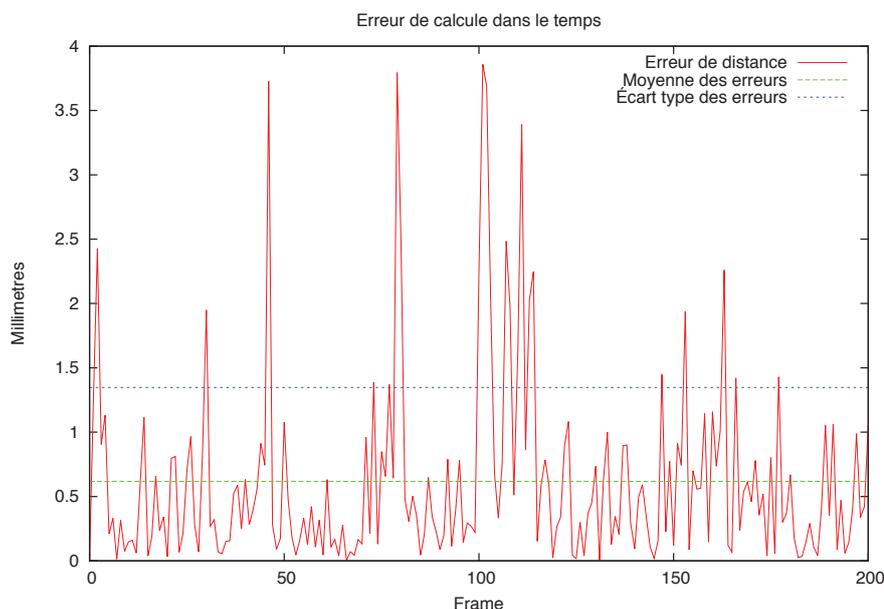


FIGURE 7.6: Exemple de l'évolution dans le temps de l'erreur de distance entre la pose simulée et la pose calculée.

Le tableau 7.1 présente des résultats du calcul de la précision de la méthode, et il conforte les résultats présentés à la Fig. 7.6. Les résultats présentés dans ce tableau sont conformes aux attentes des radiophysiciens du Centre Léon Bérard, notamment au niveau des rotations.

TABLE 7.1: Résultats de l'évaluation de la précision de la méthode de calcul de la pose sur des images synthétiques (erreur en translation et en rotation).

	Valeur moyenne	Écart type	Erreur maximum calculée	Valeur maximale simulée
Accélérateur (deg)	0.2620	0.6502	2.8720	5
Accélérateur (mm)	9.3021	5.6849	36.4512	20
Table (deg)	0.1216	0.4174	2.0975	5
Table (mm)	7.2962	6.1743	38.5880	20

7.2.2 Résultats des tests sur des données réelles

L'exécution de la méthode de suivi de mouvement sur des données réelles nous a permis d'évaluer son comportement face à des données contenant du bruit, tant au niveau des images qu'au niveau de la description de la scène (des éléments non modélisés ou des manques ou fausses informations dans le modèle numérique d'un élément).

Nous avons utilisé des webcams *Logitech* haut de gamme, et réalisé des acquisitions avec une résolution en pixels de 800×600 avec une fréquence d'acquisition de 20 images par seconde. Le système est mis en place dans la salle de traitement comme présenté à la section 6.5. Pour faciliter les tests, les images acquises sont enregistrées et traitées *a posteriori*, bien que l'objectif final soit de traiter les images pendant la séance : d'abord pour une validation clinique et ensuite pour la mise en place du système. Dans les résultats présentés, nous nous focalisons sur trois éléments : l'accélérateur, la table de traitement et un « fantôme ».

Les images de la Fig. 7.7 sont une première illustration des résultats sur une scène réelle. À chaque instant t , les informations extraites des images acquises permettent de mettre à jour le modèle à travers la méthode de suivi de mouvement développée. Une vue en réalité augmentée atteste de la bonne corrélation entre les environnements réel et virtuel par rapport à la corrélation des images réelle et virtuelle.

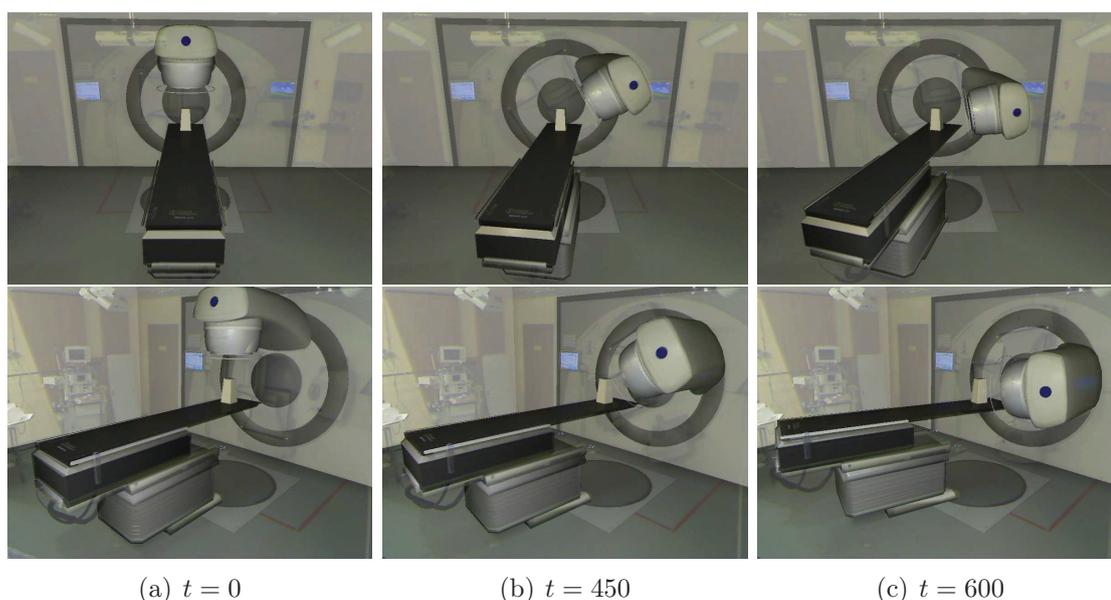
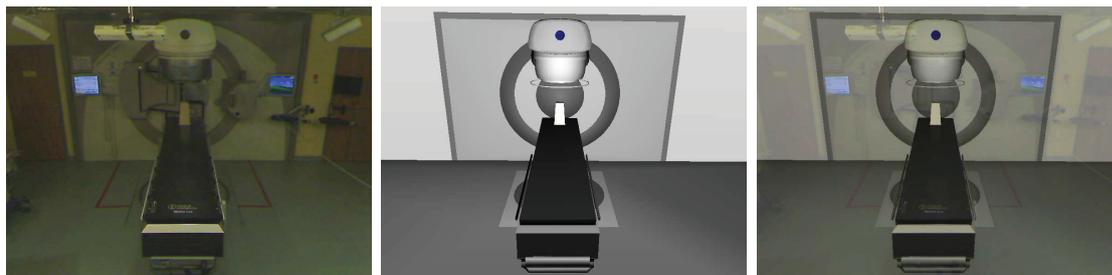
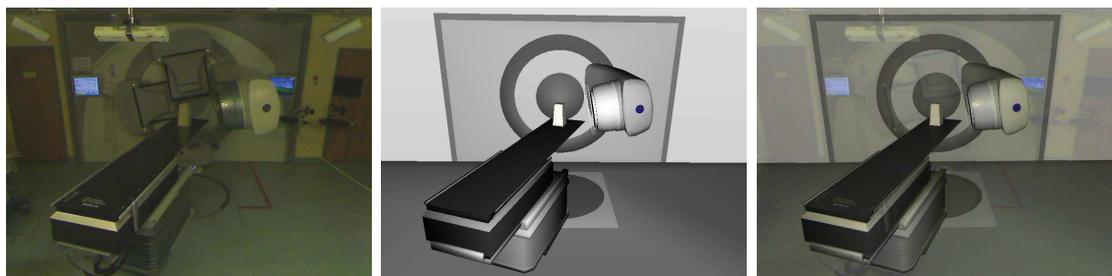


FIGURE 7.7: Résultats visuels à différents instant t en vue en réalité augmentée. Une ligne par caméra. Gauche : images réelles ; centre : images virtuelles ; droite : images en réalité augmentée.

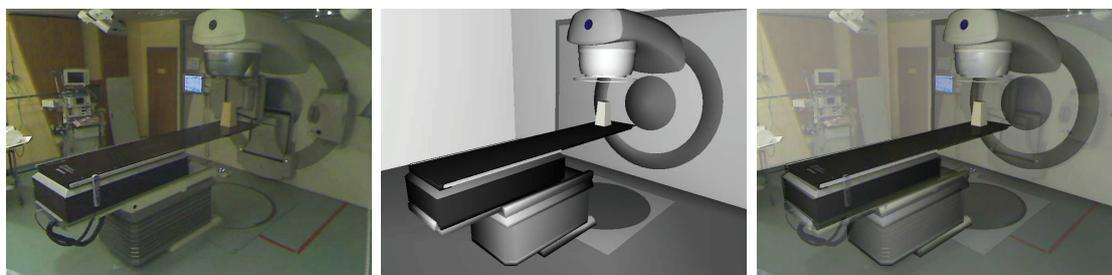
Dans les images de la Fig. 7.8, nous présentons des images des trois environnements (réel, virtuel et réalité augmentée) afin de mieux illustrer les résultats. Une procédure initialisation manuelle permet d'obtenir une corrélation conforme des environnements réel et virtuel à l'instant $t = 0$ (cf. section 3.3.6). Lorsque le mouvement d'un élément est détecté dans une image réelle acquise, la méthode de suivi de mouvement est exécutée et le modèle est mis à jour avec les informations obtenues. La superposition d'une image virtuelle, générée après la mise à jour du modèle, sur



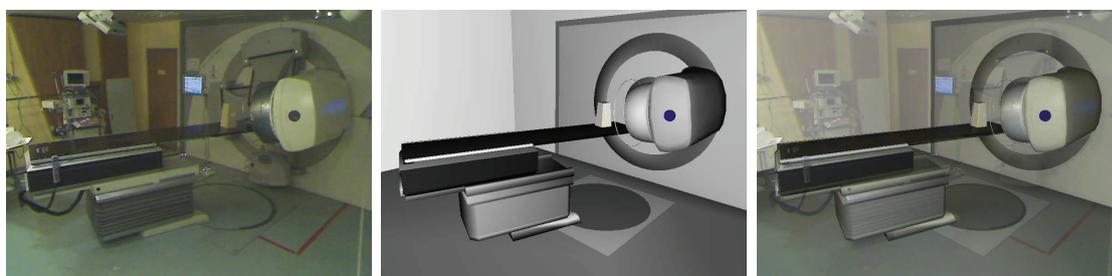
(a) Caméra 1 - Images à l'instant $t = 0$



(b) Caméra 1 - Images à un instant $t \neq 0$



(c) Caméra 2 - Images à l'instant $t = 0$



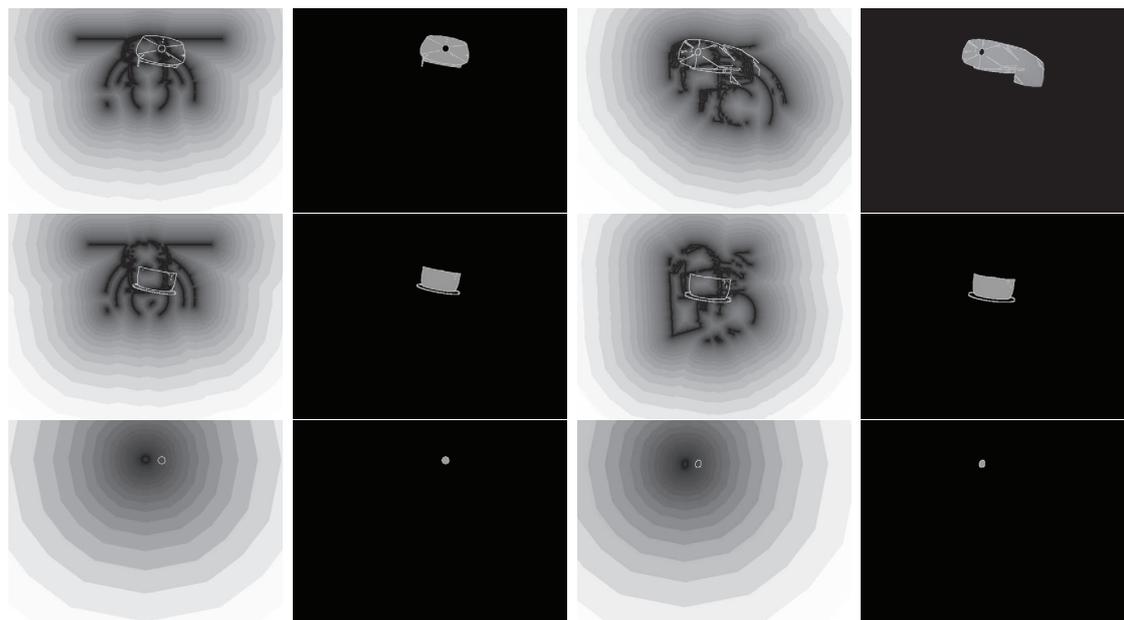
(d) Caméra 2 - Images à un instant $t \neq 0$

FIGURE 7.8: Résultats visuels à différents instant t . Gauche : images réelles ; centre : images virtuelles ; droite : images en réalité augmentée.

une image réelle permet de présenter un témoin visuel de la corrélation du modèle numérique par rapport à la scène réelle.

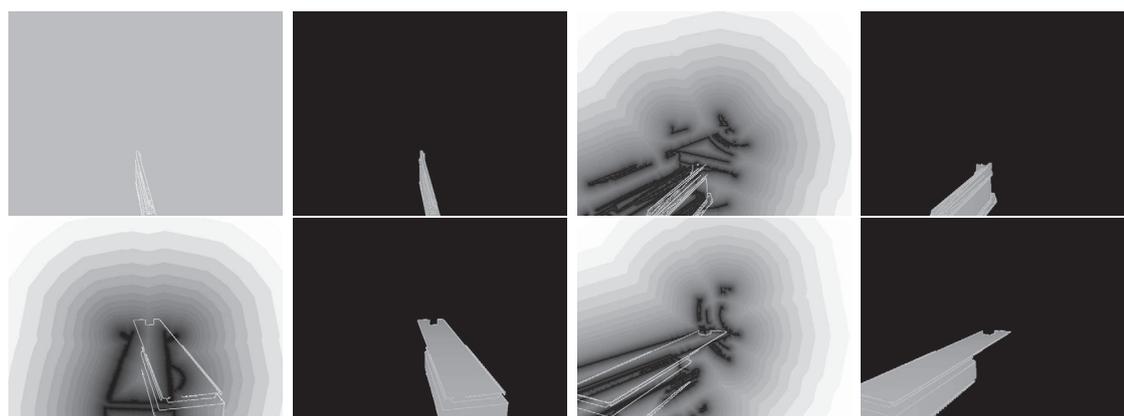
En allant plus dans les détails de l'exécution de la méthode sur le calcul de la pose d'un élément à un instant t , nous présentons les images illustrant les données traitées par la méthode à la Fig. 7.9. Dans ce cas spécifique, nous sommes dans la première itération de la méthode, ce qui veut dire que le modèle n'a pas encore été mis à jour et que les éléments ont encore la pose conforme à l'instant $t - 1$. L'intérêt de ces images est d'illustrer comment la notion de descripteur, présentée à la section 3.3.4, permet à la méthode d'être plus précise et plus robuste face aux

problèmes d'occultation.



carte de distance modèle numérique carte de distance modèle numérique

(a) Accélérateur : trois descripteurs choisis en fonction des régions de couleur différentes



carte de distance modèle numérique carte de distance modèle numérique

(b) Table de traitement : deux descripteurs choisis en fonction des degrés de liberté

FIGURE 7.9: Résultats visuels dans la première itération de la méthode de suivi de mouvement à un instant t . Une ligne par descripteur de l'élément. Les deux colonnes à gauche correspondent à la première caméra et les deux colonnes à droite à la deuxième caméra. Les colonnes impaires correspondent aux cartes de distance auxquelles nous avons superposé une image virtuelle contenant uniquement les contours du descripteur. Les colonnes paires correspondent à une image virtuelle contenant uniquement le descripteur.

C'est notamment au niveau des images contenant les cartes de distance (cf. section 4.4.1) qu'il faut se concentrer. Nous avons superposé les contours du modèle numérique du descripteur sur les cartes de distances afin de bien illustrer la différence entre les informations disponibles (carte de distances générée à partir images réelles segmentées) et les informations recherchées. Dans le cas spécifique à l'accélérateur (Fig. 7.9(a)), on peut s'apercevoir que les données pour les deux premiers

descripteurs contiennent beaucoup de bruit. Ce bruit est dû à la similarité dans la couleur entre ces deux descripteurs et les couleurs présentes dans le fond, d'où la difficulté à obtenir une bonne segmentation. Le troisième descripteur, qui est le logo de la machine, a une couleur très différente des éléments environnant et des couleurs du fond, ce qui permet d'obtenir une meilleure segmentation. Ce descripteur aide à lever les ambiguïtés, bien qu'il ne permette pas à lui tout seul de calculer la pose de l'élément, ce qui montre l'importance de la pondération des informations des descripteurs (cf. section 4.5.2).

Dans le cas spécifique de la table de traitement (Fig. 7.9(b)), on peut remarquer que le premier descripteur n'est pas vu par la première caméra (en haut à gauche), et que tous les pixels de la carte de distance associée ont la valeur maximum. Donc, les calculs associés à cette image ne seront pas ajoutés au calcul de la pose, également par la pondération des informations des descripteurs qui permettent au module de rester robuste face à l'occultation.

Avec le suivi de trois éléments (l'accélérateur, la table et un fantôme), le module permet de traiter 2 images par seconde. Cette vitesse de traitement est en partie due à l'utilisation des files de priorité (cf. section 4.5.4). Pour illustrer son apport, nous allons présenter différents cas de figure dans un exemple contenant les trois éléments. Dans le cas où on n'utilise pas de files de priorité et où les éléments sont traités dans cet ordre : l'accélérateur, la table et un fantôme, le temps de calcul total est d'environ 600 millisecondes.

Avec l'utilisation d'un opérateur (dans le modèle) définissant une contrainte de « liaison » entre le fantôme et la table de traitement, on arrive à obtenir un gain en temps de calcul d'environ 25%. Cet opérateur agit comme suit : on calcule la transformation entre la pose 3D à l'instant $t - 1$ et la pose 3D à l'instant t de l'élément A (la table dans l'exemple), et cette transformation est ensuite appliquée à la pose 3D à l'instant $t - 1$ de l'élément B (le fantôme dans l'exemple). Si cet opérateur est appliqué avant le calcul de la pose du fantôme, nous aurons déjà obtenu la pose du fantôme si celui-ci ne s'est pas déplacé par rapport à la table. En effet, lors du traitement des éléments par des files de priorités, l'évaluation de la nouvelle pose (après application de l'opérateur) permet de considérer que cette pose est déjà une bonne approximation.

En utilisant toujours le même opérateur, mais en inversant l'ordre dans lequel les éléments sont traités dans la file de priorité, le gain en temps de calcul est seulement d'environ 10%. Le fantôme est mis de côté à la première itération, et traité à l'itération suivante, sachant que la pose de la table est généralement calculée à la première itération. Ceci est dû notamment au fait que les caractéristiques de la table sont plus représentatives que celles du fantôme.

Nous travaillons actuellement sur le suivi d'individus en utilisant la méthode de suivi de mouvement proposée et le modèle générique pour un individu présenté à la section 6.5 (Fig. 7.10). Comme dans les autres cas, l'évolution du modèle numérique permet d'obtenir de meilleurs résultats. Cependant, dans le cas spécifique au patient, nous avons des difficultés supplémentaires dues principalement aux vêtements, qui ne sont pas modélisées, et à la complexité dans le nombre des degrés de liberté¹. L'une des pistes suivies actuellement consiste à faire un compromis dans la précision

1. L'humain a un total de 244 degrés de liberté dans ses articulations.

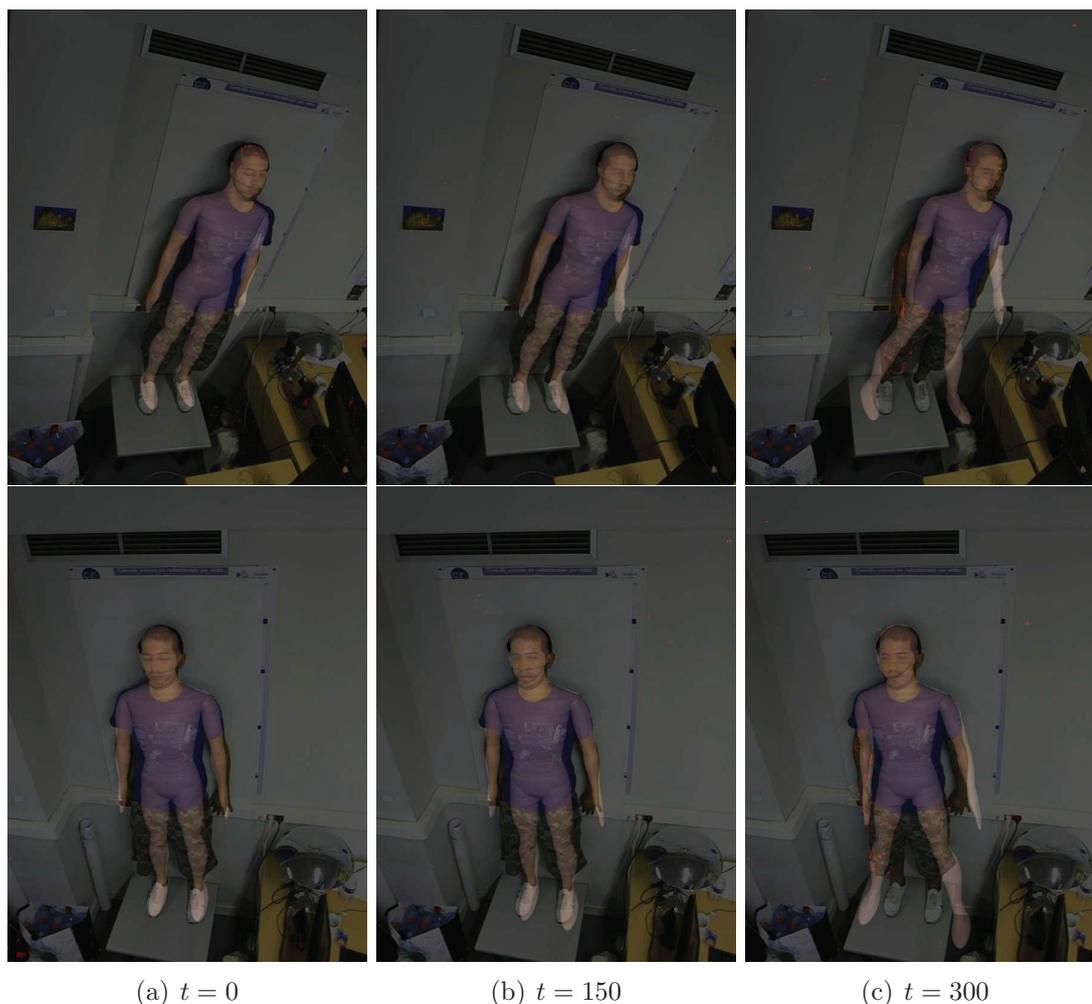


FIGURE 7.10: Résultats visuels à différents instant t en vue en réalité augmentée : utilisation du modèle générique pour un individu. Une ligne par caméra. Gauche : images réelles ; centre : images virtuelles ; droite : images en réalité augmentée.

du suivi de mouvement basé sur l'analyse temporelle : on ne s'intéresse vraiment dans la position précise du patient que quand il est allongé sur la table de traitement. Lorsque ce n'est pas le cas, un suivi moins précis de l'individu pourrait suffire.

7.2.3 Conclusions sur la méthode de suivi de mouvement

La méthode suivi de mouvement implémentée a fourni des résultats satisfaisants, avec comme principal point à améliorer le temps de calcul. Bien que nous ayons plus travaillé sur les problématiques liées à la précision et à la robustesse de la méthode, les choix des techniques utilisées ont été également faits en tenant compte de la problématique du calcul en temps interactif (10 images par seconde), voire en temps réel dans l'idéal. En effet, dès la conception de la méthode, nous avons identifié les techniques pouvant avoir une version en calcul parallèle (par exemple, la méthode du simplexe pour l'optimisation, ou le calcul en parallèle de la pose de deux éléments « indépendants », entre autres) et les techniques pouvant tirer parti de la programmation graphique. Pour ces dernières, l'idée générale serait de faire tous

les traitements des images sur carte graphique, principalement parce qu'on génère, pour chaque image réelle acquise à chaque instant t , une grande quantité d'images virtuelles : une pour chaque descripteur de chaque élément pour chaque pose 3D évaluée.

Lors des tests, nous avons pu remarquer la dépendance au modèle numérique de la méthode de suivi de mouvement, notamment au niveau de la précision des résultats. En effet, plus les informations dans le modèle d'un élément sont plus fines, plus les résultats du calcul de la pose de cet élément sont améliorés.

Cette partie du système fera l'objet d'une publication dans la revue IEEE Transactions on Biomedical Engineering (TBME) : [PDM⁺10] (actuellement en dernière révision : *minor revision*).

7.3 Reconnaissance et analyse d'événements

Pour les tests portant sur les événements, nous avons également utilisé des données simulées et des données réelles. Comme cette partie a beaucoup évolué parallèlement aux autres, grâce aux travaux d'Anthony PEREIRA (stagiaire présent sur les derniers six mois), la plupart des tests ont été réalisés avec de données réelles et simulées mélangées. Malgré tout, certains outils n'ont pas pu être développés par manque du temps, comme l'identification d'individus ou l'acquisition automatique des informations sur l'état de l'irradiation (marche ou arrêt).

7.3.1 Résultats

Nous avons d'abord réalisé des tests ponctuels, c'est-à-dire qu'on s'est intéressé à la détection et la reconnaissance d'un seul et unique événement réalisé et enregistré dans la salle de traitement :

événement « déplacement » : cet événement est plutôt lié aux résultats de la détection du mouvement d'un élément. Son intérêt est qu'au niveau de la surveillance, on va différencier la nature du déplacement selon l'élément et le moment où il arrive. Donc, dans certains cas, le seul fait qu'un élément se déplace doit générer un alerte, même sans s'intéresser à l'amplitude du déplacement. Le meilleur exemple est celui du patient pendant l'irradiation. La précision dans la détection du mouvement d'un élément est plutôt liée à la différence au niveau des pixels de l'image courante par rapport au modèle du flux vidéo (cf. section 4.3.3), mais en moyenne elle est se traduit par un déplacement de 5 millimètres pour les translations et/ou 0,5 degré pour les rotations (bien que dans certains cas, le module de suivi de mouvement puisse considérer que l'élément ne s'est pas vraiment déplacé).

événement « position cible » : dans ce cas, il s'agit de déplacer l'élément jusqu'à une position prédéfinie et ensuite de le déplacer dans une autre position quelconque. Une précision par rapport à la conformité de la pose 3D a été définie : 10 millimètres pour les translations et 1 degré pour les rotations. Dans ces conditions, la précision des résultats a été de 15 millimètres pour les translations et 1 degré pour les rotations supplémentaires, autrement dit,

on considère que la pose courante de l'élément est conforme à la pose cible en dessous d'une différence (sur la vérité terrain) de 25 millimètres pour les translations et 2 degrés pour les rotations.

événements « interaction » et « collision » : nous n'avons pas eu le temps de mettre en place les volumes d'interaction présentés à la section 5.2. Cependant, des outils contenus dans la bibliothèque logicielle Coin3D [23] permettent d'obtenir la distance minimale entre deux maillages (modèle géométrique). Ces événements demandent l'exécution d'un opérateur défini dans le modèle, chargé de calculer cette distance, et interprété selon les éléments considérés et le seuil de tolérance : on considère qu'un individu peut interagir avec n'importe quel autre élément de la salle s'il est assez proche, que deux objets peuvent interagir s'ils sont assez proches et qu'au moins l'un d'eux est en interaction avec un individu en même temps, et que des éléments risquent d'avoir une collision dans tous les autres cas. Dans les deux cas, il s'agit de détecter l'action avant qu'elle n'ait lieu, notamment pour la collision. La précision de la détection de ces événements est autour de 20 millimètres (distance entre maillages). La Fig. 7.11 illustre les manipulations réalisées pour les tests de collision d'éléments.

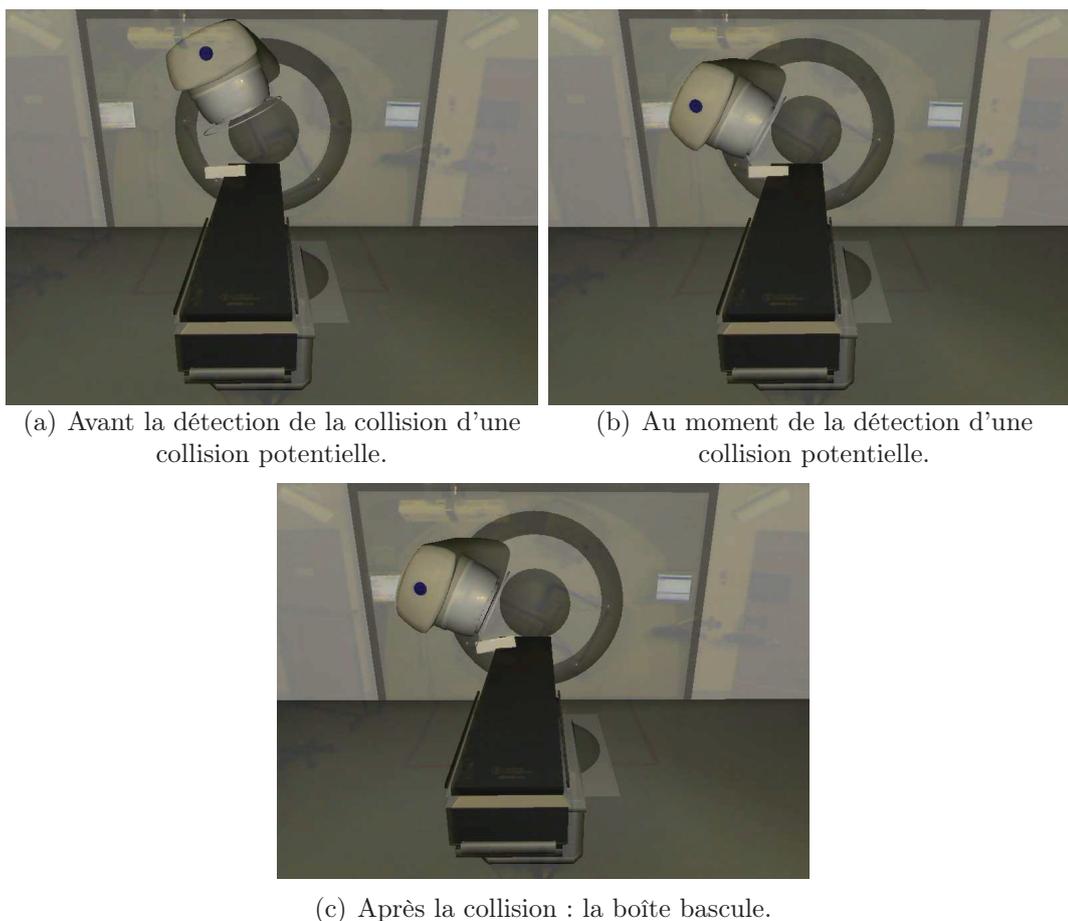


FIGURE 7.11: Résultats visuels des tests pour la détection de la collision potentielle entre l'accélérateur et un objet placé sur la table de traitement.

Ensuite, de tests sur des suites d'événements ont été réalisés pour évaluer le comportement de la méthode présentée à la section 5.3.2. À ce niveau, des événements simulés ont été rajoutés :

événement « irradiation » : cet événement est simplement ajouté à des instants bien définis pour simuler le démarrage et l'arrêt de l'irradiation. L'avantage est que le choix des instants auxquels on ajoute l'événement permet d'obtenir une grande variété de situations possibles.

événements « présence » et « identification » : notamment pour les individus. Nous n'avons pas utilisé le modèle générique d'un individu présenté à la section 6.5, mais nous avons exploité le modèle du flux vidéo (cf. section 4.3.2) et des outils de détection de blobs². Aux blobs qui ne correspondaient pas aux éléments modélisés (discrétisation faite à l'aide de masques visibilité des éléments) on ajoute des informations manuellement sur leur identification. Sur ce principe, on s'intéresse uniquement à la présence et non à la position de l'élément, qui aurait besoin d'une reconstruction 3D à partir des blobs.

On a créé des suites d'événements (scénarios) avec les événements présentés précédemment, et testé notre méthode d'analyse d'événements sur ces données. Des illustrations de manipulations réalisées pour ces tests sont présentées à la Fig. 7.12. Les résultats sont satisfaisants, avec un taux de détection de 100% de l'état de la scène (succès, autorisation, conforme, attention et danger) comme défini lors de la construction des séquences.

7.3.2 Conclusions sur le traitement d'événements

Bien que ce sous-module ne soit pas plus important que les autres, ses résultats permettent de démontrer l'intérêt et les capacités d'un système de vidéosurveillance dédié aux traitements en radiothérapie. Les résultats obtenus sont très encourageants, mais nous sommes aussi très confiants sur le fait que l'utilisation des techniques de reconnaissance de scénario vont permettre d'obtenir des informations très utiles, notamment au niveau de la prédiction d'événements et ainsi réussir éviter que des erreurs ne se produisent.

2. ensemble de pixels connectés topologiquement dans une image binaire.

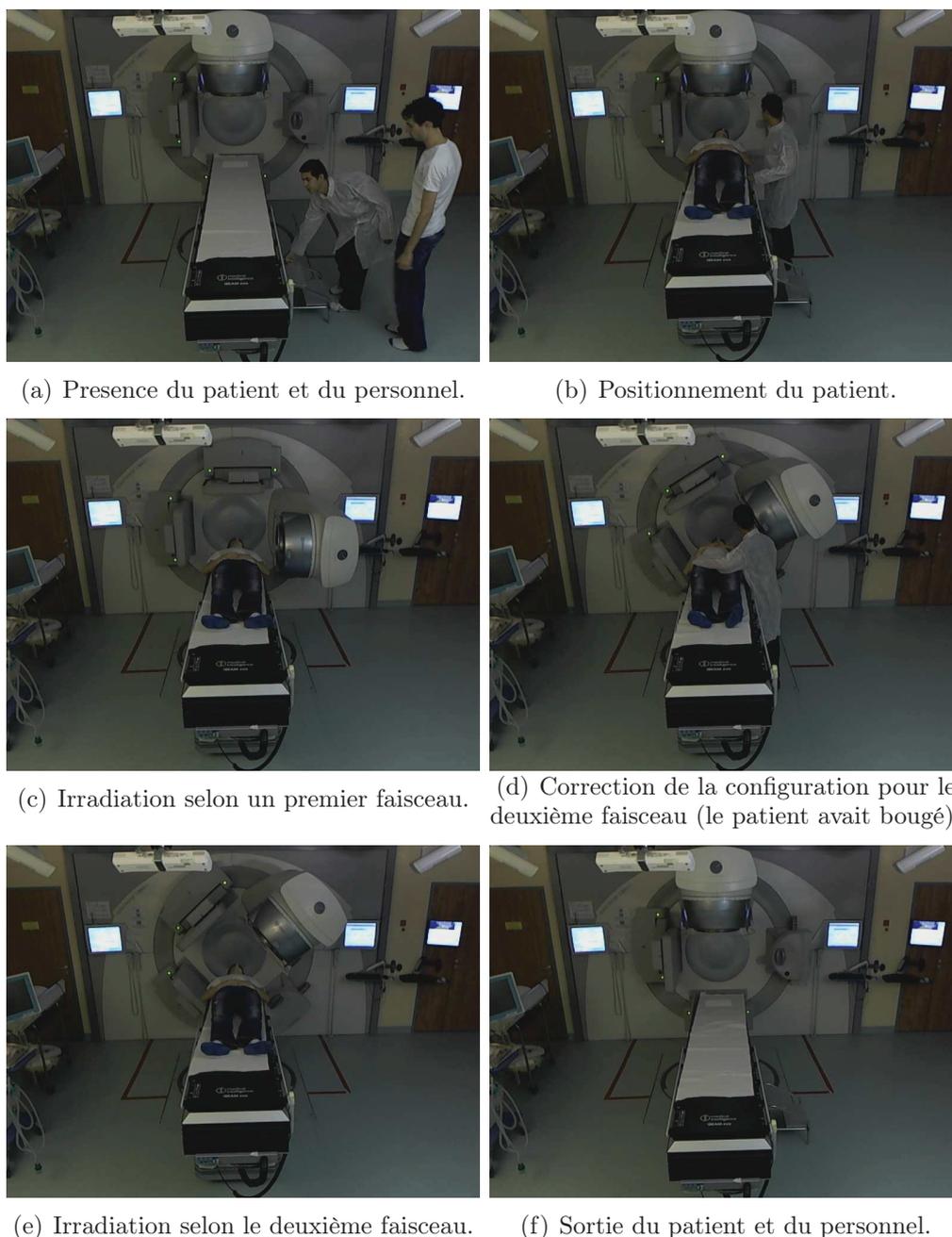


FIGURE 7.12: Exemples visuels d'une suite de situations dans le déroulement d'une séance.

Chapitre 8

Conclusion générale

Les travaux de recherche et de développement réalisés dans le cadre de cette thèse ont permis la mise en place d'un système de vidéosurveillance pour des salles de traitement en radiothérapie. Ce système est composé de plusieurs sous-modules qui ont été présentés tout au long du document, et dont les plus importants ont été détaillés. Les résultats présentés témoignent des capacités du système à contribuer à la réduction des accidents lors d'une séance de traitement en radiothérapie. Nos travaux ont profité de l'expertise des radiophysiciens du Centre Léon Bérard, et grâce au partenariat entre ce Centre et les autres acteurs de ce projet (DOSIsoft et LIRIS principalement), nous avons pu réaliser des tests dans des vraies salles de traitement avec des conditions simulées. L'état actuel du système devrait permettre de tester prochainement notre module sur des situations réelles (vraies séances de traitement filmées).

Des vingt-et-une problématiques que nous avons présentées à la section 1.3.1 nous proposons des solutions efficaces, implémentées et testées, pour quatorze d'entre elles. Notre modèle numérique générique répond aux problématiques sur la modélisation, et notre module de suivi de mouvement permet de répondre aux problématiques de vision par ordinateur et à la problématique d'optimisation de l'exploitation des informations *a priori* (*O.1*), à l'aide du modèle générique. Coté surveillance, notre module de reconnaissance d'événements couvre la problématique de détection et reconnaissance d'événements (*S.1*). Le module de traçabilité répond la problématique (*G.4*), et la modélisation d'une séance de traitement, présenté à la section 5.2, couvre la problématique (*AQ.1*). Les sept problématiques restantes sont couvertes partiellement. Notre module de reconnaissance de scénarios, une fois qu'il sera finalisé et testé dans la pratique, permettra de répondre à la problématique de détection et reconnaissance de scénarios (*S.2*). Avec l'appui d'un stagiaire du M2 d'Informatique Images de Lyon 1 financé par DOSIsoft sur les six derniers mois, nous avons réussi à formaliser cette solution, et elle est prête à être implémentée dans une suite possible de ces travaux. Les problématiques générales et en assurance qualité en radiothérapie sont couvertes partiellement (quasiment totalement) par le système de vidéosurveillance, composé de l'ensemble des méthodes proposées dans ce document. En effet, l'évolution du système permettra de couvrir entièrement ces problématiques, avec l'intégration du module de reconnaissance de scénario, des évolutions techniques visant à obtenir des résultats en temps réel notamment, et une validation clinique (tests avec des vrais patient et dans différentes services de radiothérapie).

Notre système de vidéosurveillance est le résultat d'une combinaison de techniques déjà existantes, qui ont été optimisées ou personnalisées dans le but de répondre aux besoins en assurance qualité en radiothérapie. De ce fait, nous n'avons pas réalisé de comparaison des résultats de nos méthodes avec les résultats d'autres méthodes présentées dans la littérature, car à notre connaissance il n'y a pas actuellement de méthode qui réponde aux mêmes besoins, spécifiques à une étape du processus et en même temps aux besoins globaux. Selon nous, faire une comparaison uniquement par rapport à une étape du processus n'a pas vraiment d'intérêt, et réaliser une comparaison globale sera seulement possible dès l'arrivée de systèmes équivalents au nôtre.

Un première partie du processus d'industrialisation a été assurée par Anthony PEREIRA à la fin de son stage, principalement sur les sous-modules sur lesquels il a travaillé : Interface Homme Machine et traçabilité. L'intégration des autres sous-modules est en cours d'élaboration.

8.1 Perspectives

Du fait des capacités des machines récentes (processeurs multicœurs et cartes graphiques) et du matériel d'acquisition vidéo, les avancées en informatique graphique font penser qu'on disposera prochainement de méthodes de surveillance vidéo plus précises. Ceci nous encourage à nous investir dans le développement d'applications utilisant des systèmes de suivi optique. L'une des principales perspectives est l'intégration d'un sous-module d'identification d'individus, principalement focalisé sur l'identification du patient. Un sujet de stage de fin d'étude est proposé cette année autour de ce sujet, dans le cadre de la formation de Master Recherche en Informatique option Image à l'université Lyon 1 (UCBL).

Comme précisé précédemment, une deuxième perspective très importante est l'évolution des méthodes présentées dans ce document. Cette évolution sera principalement technique, avec un but notamment industriel. Il s'agit donc d'améliorer l'implémentation de méthodes, en utilisant de la programmation parallèle ou par carte graphique, afin d'améliorer le temps de traitement des données. Une amélioration doit également être apportée à la description de la salle, car on a pu observer, lors de tests, qu'une description plus fine permet d'obtenir de meilleurs résultats. La mise en place d'une bibliothèque d'objets, au niveau du modèle numérique, permettra la réutilisation des informations et le personnalisation par l'utilisateur du modèle de la salle (et non par la personne mettant en place le système, comme cela est fait actuellement). Il est également prévu que le système puisse proposer le « revisionnage » de la trace d'une ou plusieurs séances, dans le but de permettre au personnel médical de revoir le déroulement des séances précédentes pour un patient donné. Dans cette optique, il est important d'enrichir la trace du déroulement d'une séance afin que les informations enregistrées soient suffisantes pour un analyse *a posteriori* par le personnel médical. Enfin, une étude plus approfondie sur le positionnement des caméras dans la salle devrait être réalisée.

Les méthodes proposées, dans le cadre de cette thèse, sur la vidéosurveillance d'un salle de traitement par radiothérapie seront reprises dans une suite probable, soit sur la forme d'un post-doctorat, soit dans le cadre d'un projet de Laboratoires

d'Excellence (LABEX), dont le laboratoire LIRIS est impliqué et le centre ETOILE comme la société DOSIsoft sont des partenaires, et qui présente un axe de recherche incluant la sécurisation du patient et le « patient virtuel ».

Les progrès en informatique, physique et robotique vont dans le sens de l'assistance, et parfois de l'automatisation, des traitements médicaux par des systèmes artificiels. Dans cette optique, de nouveaux équipements sont intégrés tous les jours dans les salles de traitement dans le monde, et pas uniquement en radiothérapie. Ces équipements ont des configurations spécifiques à respecter, comme c'est le cas des équipements en radiothérapie. Bien qu'en dehors des traitements par radiothérapie, les scénarios soient bien plus complexes, notamment parce que le personnel médical est présent tout au long des traitements, nous sommes confiants que des systèmes comme le notre pourraient être facilement appliqués à ces situations cliniques.

Bibliographie

- [AAK71] Y. I. Abdel-Aziz and H. M. Karara. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the Symposium on Close-Range photogrammetry*, volume 1, page 18, 1971.
- [ADM03] A. Amer, E. Dubois, and A. Mitiche. Real-time system for high-level video representation : application to video surveillance. In *Proc. SPIE Int. Symposium on Electronic Imaging, Conf. on Visual Communication and Image Processing (VCIP)*, volume 5022, pages 530–541, Santa Clara, USA, Janvier 2003.
- [AF10] Z. Arican and P. Frossard. OmniSIFT : Scale Invariant Features in Omnidirectional Images. In *Proc. of ICIP 2010*, pages 405–418, Hong Kong, Septembre 2010.
- [AvKLM01] Maysam F. Abbod, Diedrich G. von Keyserlingk, Derek A. Linkens, and Mahdi Mahfouf. Survey of utilisation of fuzzy technology in Medicine and Healthcare. *Fuzzy Sets and Systems*, 120(2) :331–349, 2001.
- [Azu97] Ronald T. Azuma. A survey of augmented reality. *Presence : Teleoperators and Virtual Environments*, 6(4) :355–385, Août 1997.
- [BBA⁺07] Bernd Bickel, Mario Botsch, Roland Angst, Wojciech Matusik, Miguel Otaduy, Hanspeter Pfister, and Markus Gross. Multi-scale capture of facial geometry and motion. *ACM Trans. Graph.*, 26, Juillet 2007.
- [BBE⁺08] A. Baumann, M. Boltz, J. Ebling, M. Koenig, H.S. Loos, M. Merkel, W. Niem, J.K. Warzelhan, and J. Yu. A Review and Comparison of Measures for Automatic Video Surveillance Systems. 2008.
- [BCP⁺93] James M. Balter, George T.Y. Chen, Charles A. Pelizzari, Swarna Krishnasamy, Steven Rubin, and Srinivasan Vijayakumar. Online repositioning during treatment of the prostate : A study of potential limits and gains. *International Journal of Radiation Oncology*Biography*Physics*, 27(1) :137–143, 1993.
- [BFO92] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world : seeing ultrasound imagery within the patient. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques, SIGGRAPH '92*, pages 203–210, New York, USA, 1992. ACM.

- [BFOP00] Guido Baroni, Giancarlo Ferrigno, Roberto Orecchia, and Antonio Pedotti. Real-time three-dimensional motion analysis for patient positioning verification. *Radiotherapy and Oncology*, 54(1) :21–27, Janvier 2000.
- [BG08] Er Bogomjakov and Craig Gotsman. Reduced Depth and Visual Hulls of Complex 3D Scenes. In 27, editor, *Computer Graphics Forum*, volume 2, pages 175—182, Avril 2008.
- [BGS07] Keni Bernardin, Tobias Gehrig, and Rainer Stiefelhagen. Multi- and single view multiperson tracking for smart room environments. In *Proceedings of the 1st international evaluation conference on Classification of events, activities and relationships*, CLEAR’06, pages 81–92, Berlin, Heidelberg, 2007. Springer-Verlag.
- [BHB10] Ouissem Ben Henia, Mohamed Hariti, and Saida Bouakaz. A two-step minimization algorithm for model-based hand tracking. In *18th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG)*, Février 2010.
- [BJD11] L. Bouarfa, PP. Jonker, and J. Dankelman. Discovery of high-level tasks in the operating room. *J Biomed Inform*, 44(3) :455–462, Juin 2011.
- [BJS⁺04] Ross I Berbeco, Steve B Jiang, Gregory C Sharp, George T Y Chen, Hassan Mostafavi, and Hiroki Shirato. Integrated radiotherapy imaging system (IRIS) : design considerations of tumour tracking with linac gantry-mounted diagnostic x-ray systems with flat-panel detectors. *Physics in Medicine and Biology*, 49(2) :243, 2004.
- [BKMM⁺04] Matthieu Bray, Esther Koller-Meier, Pascal Müller, Luc Van Gool, and Nicol N. Schraudolph. 3D Hand Tracking by Rapid Stochastic Gradient Descent Using a Skinning Model. In *First European Conference on Visual Media Production (CVMP)*, pages 59–68, London, 2004.
- [BLS⁺95] James M. Balter, Kwok L. Lam, Howard M. Sandler, J. Fred Littles, Robert L. Bree, and Randall K. Ten Haken. Automated localization of the prostate at the time of treatment using implanted radiopaque markers : Technical feasibility. *International Journal of Radiation Oncology*Biography*Physics*, 33(5) :1281–1286, 1995. Implementation of Three Dimensional Conformal Radiotherapy.
- [BM98] Francois Bremond and Gerard Medioni. *Scenario Recognition in Airborne Video Imagery*, pages 211–216. 1998.
- [BMC08] Pished Bunnun and Walterio W. Mayol-Cuevas. OutlinAR : an assisted interactive model building system with reduced computational effort. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR ’08, pages 61–64, Washington, DC, USA, 2008. IEEE Computer Society.
- [BMDT09] M. Baumann, P. Mozer, V. Daanen, and J. Troccaz. Transrectal ultrasound prostate biopsy tracking with efficient and accurate deformation estimation. In *IEEE International Symposium on Biomedical Imaging, ISBI’09*, pages 1211–1214, Boston, Juin 2009.

- [BMP04] Christoph Bregler, Jitendra Malik, and Katherine Pullen. Twist Based Acquisition and Tracking of Animal and Human Kinematics. *Int. J. Comput. Vision*, 56(3) :179–194, 2004.
- [Bor86] Gunilla Borgefors. Distance transformations in digital images. *Comput. Vision Graph. Image Process.*, 34 :344–371, Juin 1986.
- [Bou00] Jean-Yves Bouguet. Pyramidal Implementation of the Lucas Kanade Feature Tracker - Description of the algorithm, 2000.
- [BR03] Christian Blum and Andrea Roli. Metaheuristics in combinatorial optimization : Overview and conceptual comparison. *ACM Comput. Surv.*, 35 :268–308, Septembre 2003.
- [Bra98] Gary R. Bradski". Computer Vision Face Tracking For Use in a Perceptual User Interface. *Intel Technology Journal*, (Q2) :15, 1998.
- [Bro88] G. Brogefors. Hierarchical Chamfer Matching : A Parametric Edge Matching Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10 :849–865, 1988.
- [Bro92] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4) :325–376, 1992.
- [BTVG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF : Speeded Up Robust Features. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*, pages 404–417. Springer Berlin / Heidelberg, 2006.
- [BWN⁺05] James M. Balter, J. Nelson Wright, Laurence J. Newell, Barry Friemel, Steven Dimmer, Yuki Cheng, John Wong, Edward Vertatschitsch, and Timothy P. Mate. Accuracy of a wireless localization system for radiotherapy. *International Journal of Radiation Oncology*Biography*Physics*, 61(3) :933–937, 2005.
- [BZSR09] Romain Buttin, Florence Zara, Behzad Shariat, and Tanneguy Redarce. A Biomechanical Model of the Female Reproductive System and the Fetus for the realization of a Childbirth Virtual Simulator. In *IEEE Engineering in Medicine and Biology Society (EMBC'09)*, Septembre 2009.
- [CBP⁺10] Brenda G. Clark, Robert J. Brown, Jodi L. Ploquin, Anneke L. Kind, and Laval Grimard. The management of radiation treatment error through incident learning. *Radiotherapy and Oncology*, 95(3) :344–349, Juin 2010.
- [CBV⁺75] W.G. Connor, M.L.M. Boone, R. Veomett, J. Hicks, R.C. Miller, E. Mayer, and N. Sheeley. Patient repositioning and motion detection using a video cancellation system. *International Journal of Radiation Oncology Biology Physics*, 1(1-2) :147–153, Octobre-Novembre 1975.
- [CC03] Amit K. Roy Chowdhury and Rama Chellappa. A Factorization Approach for Activity Recognition. *Computer Vision and Pattern Recognition Workshop*, 4 :41, Juin 2003.

- [CCKH10] Joanne Cunningham, Mary Coffey, Tommy Knöös, and Ola Holmberg. Radiation Oncology Safety Information System (ROSIS) – Profiles of participants and the first 1074 incident reports. *Radiotherapy and Oncology*, 97(3) :601–607, Décembre 2010.
- [CDA99] Stéphane Cotin, Hervé Delingette, and Nicholas Ayache. Real-Time Elastic Deformations of Soft Tissues for Surgery Simulation. *IEEE Transactions on Visualization and Computer Graphics*, 5(1), Janvier-Mars 1999.
- [CGIP01] Antonio Chella, Maria Donatella Guarino, Ignazio Infantino, and Roberto Pirrone. A Vision System for Symbolic Interpretation of Dynamic Scenes Using ARSOM. *Applied Artificial Intelligence*, 15(8) :723–734, 2001.
- [CIR⁺10] Alvita J. Chan, Mohammad K. Islam, Tara Rosewall, David A. Jaffray, Anthony C. Easty, and Joseph A. Cafazzo. The use of human factors methods to identify and mitigate safety issues in radiation therapy. *Radiotherapy and Oncology*, 97(3) :596–600, Décembre 2010.
- [CLC10] Jing-Ying Chang, Huei-Hung Liao, and Liang-Gee Chen. Localized detection of abandoned luggage. *EURASIP J. Adv. Signal Process*, 2010 :11 :1–11 :10, Février 2010.
- [CMMS05] J. Coughlan, Roberto Manduchi, M. Mutsuzaki, and H. Shen. Rapid and Robust Algorithms for Detecting Colour Targets. In *Proc. 10th Congress of the International Colour Association (AIC Colour 05)*, pages 959–62, 2005.
- [CPV07] R. Cucchiara, A. Prati, and R. Vezzani. A multi-camera vision system for fall detection and alarm generation. 24(5) :334–345, 2007.
- [CSM⁺03] Sebastien Clippe, David Sarrut, Claude Malet, Serge Miguet, Chantal Ginestet, and Christian Carrie. Patient setup error measurement using 3D intensity-based image registration techniques. *International Journal of Radiation Oncology Biology Physics*, 56(1) :259–265, Mai 2003.
- [DA94] René David and Hassane Alla. Petri nets for modeling of dynamic systems : A survey. *Automatica*, 30(2) :175–202, 1994.
- [Dan80] P.-E. Danielsson. Euclidean distance mapping. *Computer Graphics and Image Processing*, 14 :227–248, 1980.
- [DBN⁺11] F. Debouck, J. Bourhis, G. Noël, H. Petit, L. Ravinet, and T. Didier. ORION® : Simple and effective method for systemic analysis of event occurring in hospital practice. In *Acts of the 30th European Society for Therapeutic Radiology and Oncology (ESTRO)*, 2011.
- [DBR00] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 126–133, 2000.
- [DC01] Tom Drummond and Roberto Cipolla. Real-Time Tracking of Highly Articulated Structures in the Presence of Noisy Measurements. In

- In IEEE International Conference on Computer Vision (ICCV)*, pages 315–320, 2001.
- [dCL11] M. do Carmo Lopes. The role of education and training in providing safe radiotherapy. In *Acts of the 30th European Society for Therapeutic Radiology and Oncology (ESTRO)*, 2011.
- [dedeSA03] Agence Nationale d’Accréditation et d’Évaluation en Santé (ANAES). Principes méthodologiques pour la gestion des risques en établissement de santé. Technical report, Haute Autorité de Santé, France, Janvier 2003.
- [DF01] Quentin Delamarre and Olivier Faugeras. 3D Articulated Models and Multiview Tracking with Physical Forces. *Computer Vision and Image Understanding*, 81(3) :328–357, 2001.
- [DG94] Christophe Dousson and Malik Ghallab. Suivi et reconnaissance de chroniques. *Revue d’intelligence artificielle*, 8(1) :29–61, 1994.
- [DGG93] Christophe Dousson, Paul Gaborit, and Malik Ghallab. Situation recognition : representation and algorithms. In *Proceedings of the 13th international joint conference on Artificial intelligence - Volume 1*, pages 166–172, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.
- [dJNR11] R. de Jong, J. Nijkamp, and C. Rasch. First results of a worldwide survey on responsibilities and resources in modern radiation therapy. In *Acts of the 30th European Society for Therapeutic Radiology and Oncology (ESTRO)*, 2011.
- [DL05] Elodie Desserée and Louis Legrand. First Results of a Complete Marker-Free Methodology for Human Gait Analysis. In *IEEE EMBC 2005*, Septembre 2005.
- [dlSedSF09] Ministère de la Santé et des Sports (FRANCE). Arrêté du 22 janvier 2009 portant homologation de la décision no 2008-DC-0103 de l’Autorité de sûreté nucléaire du 1er juillet 2008 fixant les obligations d’assurance de la qualité en radiothérapie définies à l’article R. 1333-59 du code de la santé publique. Technical report, Ministère de la Santé et des Sports (FRANCE), France, Janvier 2009.
- [DNT01] Emmanuel Dubois, Laurence Nigay, and Jocelyne Troccaz. Consistency in Augmented Reality Systems. In Murray Little and Laurence Nigay, editors, *Engineering for Human-Computer Interaction*, volume 2254 of *Lecture Notes in Computer Science*, pages 111–122. Springer Berlin / Heidelberg, 2001.
- [DRF98] Michel Dojat, Nicolas Ramaux, and Dominique Fontaine. Scenario recognition for temporal reasoning in medical domains. *Artificial Intelligence in Medicine*, pages 139–155, 1998.
- [dSN08] ASN : Autorité de Sûreté Nucleaire. Décision n° 2008-DC-0103 de l’Autorité de sûreté nucléaire du 1er juillet 2008 fixant les obligations d’assurance de la qualité en radiothérapie. Technical report, ASN, France, Juillet 2008.

- [dSN10] ASN : Autorité de Sureté Nucleaire. Les échelles de classement des incidents et accidents nucléaires et des événements en radioprotection dans le cadre de radiothérapies. Technical report, ASN, France, Mars 2010.
- [dSN11] ASN : Autorité de Sureté Nucleaire. La securité du patient - Pour une dynamique du progrès. Technical Report 1, ASN, France, Mars 2011.
- [DT03] James W. Davis and Ambrish Tyagi. A Reliable-Inference Framework for Recognition of Human Actions. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS '03*, page 169, Washington, DC, USA, 2003. IEEE Computer Society.
- [DW05] A. Dipanda and S. Woo. Towards a real-time 3D shape reconstruction using a structured light system. *Pattern Recognition*, 38(10) :1632–1650, 2005.
- [ELMM05] Ersin Elbasi, Zuo Long, Kishan Mehrotra, and C Mohan. Control charts approach for scenario recognition in video sequences. *Turk J Elec Engin*, 13(3) :303–309, 2005.
- [FAHF08] Andrea Fossati, Élise Arnaud, Radu P. Horaud, and Pascal Fua. Tracking articulated bodies using generalized expectation maximization. In *Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA) at Computer Vision and Pattern Recognition, CVPR'08*, pages 1–6, Anchorage, Etats-Unis, Juin 2008. IEEE.
- [FFG⁺01] G. Fangi, F. Fiori, G. Gagliardini, E. Savina, and E. S. Malinverni. Fast And Accurate Close Range 3D Modelling By Laser Scanning System. In *Proc. 18th Int. Symposium of CIPA*, Postdam, Germany, 2001.
- [FGI01] R Fried, U Gather, and M Imhoff. Online pattern recognition in intensive care medicine. *Proc AMIA Symp*, pages 184–8, 2001.
- [FH04] Pedro F Felzenszwalb and Daniel P Huttenlocher. Distance Transforms of Sampled Functions. *Cornell Computing and Information Science TR20041963*, 2004.
- [Fia05] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 590–596, 2005.
- [FL09] P. Francois and E. Lartigau. Analyse des risques en radiothérapie - Risk analysis in radiotherapy. *Cancer/Radiothérapie*, 13(6-7) :574–580, 2009.
- [FNFB04] J. Fischer, M. Neff, D. Freudenstein, and D. Bartz. Medical Augmented Reality based on Commercial Image Guided Surgery. In S. Coquillart and M. Göbel, editors, *Eurographics Symposium on Virtual Environments (EGVE 2004)*, pages 83–86, 2004.
- [FS06] R. Freeman and A. Steed. Interactive modelling and tracking for mixed and augmented reality. In *Proceedings of the ACM symposium*

- on *Virtual reality software and technology*, VRST '06, pages 61–64, New York, NY, USA, 2006. ACM.
- [FSC⁺04] Martin Fuss, Bill J. Salter, Sean X. Cavanaugh, Cristina Fuss, Amir Sadeghi, Clifton D. Fuller, Ardow Ameduri, James M. Hevezi, Terence S. Herman, and Charles R. Thomas. Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies. *International Journal of Radiation Oncology*Biography*Physics*, 59(4) :1245–1256, 2004.
- [FSS⁺01] M. Fuss, B. J. Salter, A. Sadeghi, C. R. Thomas, A. Ameduri, S. Chopra, J. M. Hevezi, and T. S. Herman. Stereotactic BAT(TM) ultrasound positioning for upper abdominal target volumes. *International Journal of Radiation Oncology*Biography*Physics*, 51(3, Supplement 1) :259–259, 2001.
- [Gav99] Dariu Gavrilă. The Visual Analysis of Human Movement : A Survey. *Computer Vision and Image Understanding*, 73(1) :82–98, 1999.
- [GBCN10] Michela Goffredo, Imed Bouchrika, John N. Carter, and Mark S. Nixon. Performance analysis for automated gait extraction and recognition in multi-camera surveillance. *Multimedia Tools Appl.*, 50 :75–94, Octobre 2010.
- [GD96] D. M. Gavrilă and L. S. Davis. 3-D model-based tracking of humans in action : a multi-view approach. In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, CVPR '96, pages 73–, Washington, DC, USA, 1996. IEEE Computer Society.
- [GDM05] Oxana Govokhina, Elodie Desserée, and Jean-Michel Moreau. Reconstruction of breathing movements from videos, Janvier 2005.
- [GEW⁺96] W.E.L. Grimson, G.J. Ettinger, S.J. White, T. Lozano-Perez, W.M. Wells, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Transactions on Medical Imaging*, 15 :129–140, Avril 1996.
- [GJH09] D.A. Gomez Jauregui and P. Horain. Region-Based vs. Edge-Based Registration for 3D Motion Capture by Real Time Monoscopic Vision. In *MIRAGE09*, pages 344–355, 2009.
- [Gle99] Michael Gleicher. Animation from observation : Motion capture and motion editing. *SIGGRAPH Comput. Graph.*, 33 :51–54, Novembre 1999.
- [Glo86] Fred Glover. Future paths for integer programming and links to artificial intelligence. *Comput. Oper. Res.*, 13 :533–549, Mai 1986.
- [Gol89] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [GPC11] Giovanni Gualdi, Andrea Prati, and Rita Cucchiara. Contextual information and covariance descriptors for people surveillance : an ap-

- plication for safety of construction workers. *J. Image Video Process.*, 2011 :9 :1–9 :16, Janvier 2011.
- [GSS93] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2) :107–113, Avril 1993.
- [Gui00] *Guidelines on education and training in radiation protection for medical exposures*. Radiation protection. Office for Official Publications of the European Communities, 2000.
- [GVPG03] Pierre F. Gabriel, Jacques G. Verly, Justus H. Piater, and André Genon. The state of the art in multiple object tracking under occlusion in video sequences. In *In Advanced Concepts for Intelligent Vision Systems (ACIVS), 2003*, pages 166–173, 2003.
- [GYB04] S. Burak Gokturk, Hakan Yalcin, and Cyrus Bamji. A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions. In *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 3 - Volume 03*, page 35, Washington, DC, USA, 2004. IEEE Computer Society.
- [HB02] Patrick Horain and Mayank Bomb. 3D Model Based Gesture Acquisition Using a Single Camera. In *WACV '02 : Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, page 158, Washington, DC, USA, 2002. IEEE Computer Society.
- [HBN00a] Somboon Hongeng, Francois Bremond, and Ramakant Nevatia. Bayesian Framework for Video Surveillance Application. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, page 1164, Washington, DC, USA, 2000. IEEE Computer Society.
- [HBN00b] Somboon Hongeng, Francois Bremond, and Ramakant Nevatia. Representation and Optimal Recognition of Human Activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, Juin 2000.
- [HCP01] Carine Hue, Jean-Pierre Le Cadre, and Patrick Pérez. A Particle Filter to Track Multiple Objects. In *Proceedings of the IEEE Workshop on Multi-Object Tracking (WOMOT'01)*, page 61, Washington, DC, USA, 2001. IEEE Computer Society.
- [HHC96] Seth Hutchinson, Greg Hager, and Peter Corke. A Tutorial on Visual Servo Control. *IEEE Transactions on Robotics and Automation*, 12 :651–670, 1996.
- [HHD99] Ismail Haritaoglu, David Harwood, and Larry S. Davis. Hydra : Multiple People Detection and Tracking Using Silhouettes. In *Proceedings of the Second IEEE Workshop on Visual Surveillance, VS '99*, page 6, Washington, DC, USA, 1999. IEEE Computer Society.
- [HHE03] Raffay Hamid, Yan Huang, and Irfan Essa. ARGMode - Activity Recognition using Graphical Models. *Computer Vision and Pattern Recognition Workshop*, 4 :38, 2003.
- [HKKR93] Daniel P. Huttenlocher, Gregory A. Klanderman, Gregory A. Kl, and William J. Rucklidge. Comparing Images Using the Hausdorff

- Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 :850–863, 1993.
- [HME⁺05] A Hsu, N R Miller, P M Evans, J C Bamber, and S Webb. Feasibility of using ultrasound for real-time tracking during radiotherapy. *Medical Physics*, 32(6) :1500–1512, 2005.
- [HN01] S. Hongeng and R. Nevatia. Multi-agent event recognition. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 84–91, 2001.
- [HS80] Berthold K.P. Horn and Brian G. Schunck. Determining Optical Flow. Technical report, Cambridge, MA, USA, 1980.
- [HS88] C. Harris and M.J. Stephens. A Combined Corner and Edge Detector. In *4th Alvey Vision Conference*, pages 147–152, Manchester, UK, 1988.
- [HSC98] G B Hanna, S M Shimi, and A Cuschieri. Task performance in endoscopic surgery is influenced by location of the image displa. *Annals of Surgery*, 227(4) :481–484, Août 1998.
- [HTWM04] W. Hu, T.N. Tan, L. Wang, and S.J. Maybank. A Survey on Visual Surveillance of Object Motion and Behaviors. 34(3) :334–352, Août 2004.
- [HUF04] L. Herda, R. Urtasun, and P. Fua. Hierarchical Implicit Surface Joint Limits to Constrain Video-Based Motion Capture. In *Computer Vision – ECCV 2004*, volume II, pages 405–418, 2004.
- [IB98] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1) :5–28, 1998.
- [IM01] Michael Isard and John MacCormick. BraMBLe : A Bayesian Multiple-Blob Tracker. In *ICCV*, pages 34–41, 2001.
- [JMH⁺99] L.S. Johnson, B.D. Milliken, S.W. Hadley, C.A. Pelizzari, D.J. Harraf, and G.T.Y. Chen. Initial clinical experience with a video-based patient positioning system. *International Journal of Radiation Oncology Biology Physics*, 45(1) :205–213, 1999.
- [Joh73] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14 :201–211, 1973.
- [Kal60] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*, (82 (Series D)) :35–45, 1960.
- [KBP⁺00] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. Virtual object manipulation on a table-top AR environment. In *International Symposium on Augmented Reality*, pages 111–119. IEEE Computer Society, 2000.
- [KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598) :671–680, 1983.

- [KLW10] K. Kim, V. Lepetit, and W. Woo. Keyframe-based Modeling and Tracking of Multiple 3D Objects. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, 2010.
- [KM95] I.A. Kakadiaris and D. Metaxas. 3D human body model acquisition from multiple views. *Computer Vision, IEEE International Conference on*, 0 :618, 1995.
- [Kol92] Janet L. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6(1) :3–34, Mars 1992.
- [KOS⁺03] Kengo Kuriyama, Hiroshi Onishi, Naoki Sano, Takafumi Komiyama, Yoshihito Aikawa, Yoshihito Tateda, Tsutomu Araki, and Minoru Uematsu. A new irradiation unit constructed of self-moving gantry-CT and linac. *International Journal of Radiation Oncology Biology Physics*, 55(2) :428–435, Février 2003.
- [LBP⁺07] R Lang, D Bruckner, G Pratl, R Velik, and T Deutsch. Scenario Recognition in Modern Building Automation. *International Journal of Intelligent Systems and Technologies*, 4-1, No. 5 :36–44, 2007.
- [LCDD⁺08] E. Lartigau, B. Coche-Dequeant, V. Dumortier, S. Giscard, T. Lacornerie, A. Lasue, V. Cheval, V. Martel, B. Malfait, A. Fuchs, M. Pestel, M. Damman, and M. Forrest. La cellule de retour d’expérience (CREx) en radiothérapie : un outil indispensable au management de la sécurité. Une évaluation après deux ans d’activité au centre Oscar-Lambret. *Cancer/Radiothérapie*, 12(6-7) :610 – 613, 2008. 19e Congrès national de la Société Française de Radiothérapie Oncologique.
- [LDL⁺03] Darren J. Little, Lei Dong, Lawrence B. Levy, Anurag Chandra, and Deborah A. Kuban. Use of portal images and BAT ultrasonography to measure setup error and organ motion for prostate IMRT : implications for treatment margins. *International Journal of Radiation Oncology*Biology*Physics*, 56(5) :1218–1224, Août 2003.
- [LDS⁺02] Dale Litzenberg, Laura A. Dawson, Howard Sandler, Martin G. Sanda, Daniel L. McShan, Randall K. Ten Haken, Kwok L. Lam, Kristy K. Brock, and James M. Balter. Daily prostate targeting using implanted radiopaque markers. *International Journal of Radiation Oncology*Biology*Physics*, 52(3) :699–703, 2002.
- [Lea94] LL. Leape. Error in medicine. *JAMA*, 272 :1851–7, 1994.
- [Lee05] Dar-Shyang Lee. Effective Gaussian Mixture Learning for Video Background Subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27 :827–832, Mai 2005.
- [LF05] Vincent Lepetit and Pascal Fua. Monocular model-based 3D tracking of rigid objects. *Found. Trends. Comput. Graph. Vis.*, 1 :1–89, Janvier 2005.
- [LK81] Bruce D. Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of the 1981 DARPA Image Understanding Workshop*, pages 121–130, Avril 1981.

- [LMP⁺99] Joseph Lattanzi, Shawn McNeeley, Wayne Pinover, Eric Horwitz, Indra Das, Timothy E. Schultheiss, and Gerald E. Hanks. A comparison of daily CT localization to a daily ultrasound-based system in prostate cancer. *International Journal of Radiation Oncology*Biology*Physics*, 43(4) :719–725, Mars 1999.
- [LMV⁺06] A Lathuilière, A Mansouri, Y Voisin, F S Marzani, and P Gouton. Stereoscopic system for 3D reconstruction using multispectral camera and LCD projector. *Imaging Science Journal*, 54(1) :19–28, Mars 2006.
- [Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Key-points. *Int. J. Comput. Vision*, 60 :91–110, Novembre 2004.
- [LPF04] Vincent Lepetit, Julien Pilet, and Pascal Fua. Point matching as a classification problem for fast and robust object pose estimation. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition*, CVPR'04, pages 244–250, Washington, DC, USA, 2004. IEEE Computer Society.
- [LW07] Donghoon Lee and Matthew Wiswall. A Parallel Implementation of the Simplex Function Minimization Routine. *Comput. Econ.*, 30 :171–187, Septembre 2007.
- [Mar87] Etienne-Jules Marey. Le mécanisme du vol des oiseaux éclairé par la chronophotographie. *La nature : revue des sciences et de leurs applications aux arts et à l'industrie*, (757) :8–14, Décembre 1887.
- [Mar63] Donald W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11(2) :431–441, 1963.
- [MB00] John MacCormick and Andrew Blake. A Probabilistic Exclusion Principle for Tracking Multiple Objects. *Int. J. Comput. Vision*, 39 :57–71, Août 2000.
- [MBCM97] Philip McLauchlan, David Beymer, Benn Coifman, and Jitendra Mali. A Real-time Computer Vision System for Measuring Traffic Parameters. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, page 495, Washington, DC, USA, 1997. IEEE Computer Society.
- [MC02] E. Marchand and F. Chaumette. Virtual visual servoing : A framework for real-time augmented reality. In G. Drettakis and H.-P. Seidel, editors, *EUROGRAPHICS 2002 Conference Proceeding*, volume 21(3) of *Computer Graphics Forum*, pages 289–298, Saarebrücken, Germany, Septembre 2002.
- [MCL01] Franck Marzani, Elodie F. Calais, and Louis Legrand. A 3-D marker-free system for the analysis of movement disabilities - an application to the legs. *IEEE Transactions on Information Technology in Biomedicine*, 5(1) :18–26, 2001.
- [mdls88] OMS :Organisation mondiale de la santé. Quality Assurance in Radiotherapy. Technical report, WHO, Geneva, 1988.

- [mdls09] OMS :Organisation mondiale de la santé. Radiotherapy risk profile : technical manual. Technical report, WHO, Geneva, Février 2009.
- [Men99] Alberto Menache. *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 1999.
- [MG01] Thomans B. Moeslund and Erik Granum. A Survey of Computer Vision-Based Human Motion Capture. *Computer Vision and Image Understanding*, 81(3) :231–268, Mars 2001.
- [MGB05] Brice Michoud, Erwan Guillou, and Saida Bouakaz. Human model and pose Reconstruction from Multi-views. In *International Conference on Machine Intelligence (ACIDCA-ICMI)*, Novembre 2005.
- [MIB⁺09] H. Medjahed, D. Istrate, J. Boudy, B. Dorizzi, and F. Steenkeste. Environnement multimodal pour la televigilance médicale à domicile EMUTEM. In *Congrès SFTAG 2009*, Troyes, Novembre 2009.
- [MKO⁺08] T.L. McDonald, L. Ku, K.M. O'Donnell, D. Kaurin, P.J. Gagnon, C.R. Thomas, A.Y. Hung, and M. Fuss. Quantifying Respiratory-induced Prostate Motion using Continuous Real-time Tracking Technology. *International Journal of Radiation Oncology*Biophysics*, 72(1, Supplement 1) :S334–S334, 2008. Proceedings of the American Society for Therapeutic Radiology and Oncology 50th Annual Meeting, American Society for Therapeutic Radiology and Oncology 50th Annual Meeting.
- [MLDR07] Daniel Modrow, Claudio Laloni, Guenter Doemens, and Gerhard Rigoll. 3D Face Scanning Systems Based on Invisible Infrared Coded Light. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Nikos Paragios, Syeda-Mahmood Tanveer, Tao Ju, Zicheng Liu, Sabine Coquillart, Carolina Cruz-Neira, Torsten Müller, and Tom Malzbender, editors, *Advances in Visual Computing*, volume 4841 of *Lecture Notes in Computer Science*, pages 521–530. Springer Berlin / Heidelberg, 2007.
- [MLS94] Richard M. Murray, Zexiang Li, and Shankar S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC, 1 edition, Mars 1994.
- [MM05] J.J. Mazon and A. Maugis. *Techniques d'irradiation des cancers : la radiothérapie conformationnelle*. Maloine, 2005.
- [MQR03] Calvin R. Maurer, Jr., Rensheng Qi, and Vijay Raghavan. A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25 :265–270, Février 2003.
- [MRH⁺97] Barret D. Milliken, Steven J. Rubin, Russel J. Hamilton, Scott L. Johnson, and George T. Y. Chen. Performance of a video-image-subtraction-based patient positioning system. *International Journal of Radiation Oncology Biology Physics*, 38(4) :855–866, 1997.
- [MSR⁺10] Roland Mörzinger, Manolis Sardis, Igor Rosenberg, Helmut Grabner, Galina Veres, Imed Bouchrika, Marcus Thaler, Rene Schuster,

- Albert Hofmann, Georg Thallinger, Vassileios Anagnostopoulos, Dimitrios Kosmopoulos, Athanasios Voulodimos, Constantinos Lalos, Nikolaos Doulamis, Theodora Varvarigou, Rolando Palma Zelada, Ignacio Jubert Soler, Severin Stalder, Luc Van Gool, Lee Middleton, Zoheir Sabeur, Banafshe Arbab-Zavar, John Carter, and Mark Nixon. Tools for semi-automatic monitoring of industrial workflows. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, ARTEMIS '10, pages 81–86, New York, NY, USA, 2010. ACM.
- [Muy87] Eadweard Muybridge. *Animal locomotion*. University of Pennsylvania, 1887.
- [MVLD00] Franck Marzani, Yvon Voisin, Lew F.C. Lew Yan Vonn, and Alain Diou. Active Stereovision System : A Fast and Easy Calibration Method. In *Proceedings of the sixth International Conference on Control, Automation, Robotics and Vision*, Singapore, Décembre 2000.
- [MWW11] P. Malarme, D. Wikler, and N. Warzée. Intraoperative capture of surgical workflow. In *Computer Assisted Radiology and Surgery (CARS) - 12th IFCARS / SPIE / EuroPACS Joint Workshop on Surgical PACS and the Digital Operating Room, Berlin, Germany*, volume 6, pages S146–S147, Juin 2011.
- [MWWW11] Damian Mierzwinski, Dariusz Walczak, Marcin Wolski, and Marcin Wrzos. Video and Sensor Data Integration in a Service-Oriented Surveillance System. *Scientific International Journal for Parallel and Distributed Computing*, 12(1) :93–103, 2011.
- [NGG11] F. Nater, H. Grabner, and L. Van Gool. Unsupervised Workflow Discovery in Industrial Environments. In *ICCV Workshop on Visual Surveillance*, 2011. in press.
- [NKY08] Srinivasa G. Narasimhan, Sanjeev Jagannatha Koppal, , and Shuntaro Yamazaki. Temporal Dithering of Illumination for Fast Active Vision. In *Proc. European Conference on Computer Vision*, pages 830–844, Octobre 2008.
- [NM65] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4) :308–313, Janvier 1965.
- [NPS⁺09] S. Nicolau, Xavier Pennec, Luc Soler, X. Buy, A. Gangi, Nicholas Ayache, and J. Marescaux. An Augmented Reality System for Liver Thermal Ablation : Design and Evaluation on Clinical Cases. *Medical Image Analysis*, 13(3) :494–506, Juin 2009.
- [NS06] David Nister and Henrik Stewenius. Scalable Recognition with a Vocabulary Tree. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, CVPR '06*, pages 2161–2168, Washington, DC, USA, 2006. IEEE Computer Society.
- [OH99a] H. Ouhaddi and P. Horain. Modélisation 3D du geste par la vision. In *Actes 5èmes Journées d'études et d'échanges Compression et REprésentation*, pages 101–108, 1999.

- sentation des Signaux Audiovisuels (CORESA '99)*, Sophi-Antipolis, France, pages 277–283, 1999.
- [OH99b] Hocine Ouahdi and Patrick Horain. 3D Hand Gesture Tracking by Model Registration. In *Proceedings of International Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'99)*, pages 70–73, Septembre 1999.
- [OKS80] Yu-Ichi Ohta, Takeo Kanade, and Toshiyuki Sakai. Color information for region segmentation. *Computer Graphics and Image Processing*, 13(3) :222–241, 1980.
- [ORS09] Songhwai Oh, Stuart Russell, and Shankar Sastry. Markov Chain Monte Carlo Data Association for Multi-target Tracking. *IEEE Trans. Automatic Control*, 54(3) :481–497, Mars 2009.
- [oRUM93] ICRU : International Commission on Radiation Units and Measurements. Prescribing, recording and reporting photon beam therapy. Technical Report Report 50, ICRU, 1993.
- [oRUstIr99] ICRU : International Commission on Radiation Units and Measurements (supplement to ICRU report 50). Prescribing, recording and reporting photon beam therapy. Technical Report Report 62, ICRU, 1999.
- [PA05] E.B. Podgoršak and International Atomic Energy Agency. *Radiation oncology physics : a handbook for teachers and students*. STI/PUB. International Atomic Energy Agency, 2005.
- [PCMC06] Muriel Pressigout, Andrew I. Comport, Eric Marchand, and Francois Chaumette. Real-Time Markerless Tracking for Augmented Reality : The Virtual Visual Servoing Framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(4) :615–628, 2006.
- [PDM⁺10] Miguel Portela Sotelo, Elodie Desserée, Jean-Michel Moreau, Behzad Shariat, and Michaël Beuve. Model-based multiple object video tracking for treatment room supervision. Technical Report RR-LIRIS-2010-023, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon, Novembre 2010. soumis à la revue IEEE Transactions on Biomedical Engineering (TBME) : actuellement en dernière révision : minor revision.
- [PDM11] Miguel Portela Sotelo, Elodie Desserée, and Jean-Michel Moreau. Multiple object video tracking for the radiotherapy treatment room supervision : focussing on the model. In *Computer Assisted Radiology and Surgery*. Springer, Juin 2011.
- [PFB⁺03] Lennert S. Ploeger, Michel Frenay, Anja Betgen, Josien A. de Bois, Kenneth G.A. Gilhuijs, and Marcel van Herk. Application of video imaging for improvement of patient set-up. *Radiotherapy and Oncology*, 68(3) :277–284, Septembre 2003.
- [Pie03] Wojciech Pieczynski. Modèles de Markov en traitements d'images. *Traitement du Signal*, 20(3) :255–278, 2003.

-
- [PLW08] Youngmin Park, Vincent Lepetit, and Woontack Woo. Multiple 3D Object tracking for augmented reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '08*, pages 117–120, Washington, DC, USA, 2008. IEEE Computer Society.
- [PMF⁺07] S. Pinault, G. Morel, R. Ferrand, M. Auger, and C. Mabit. Using an external registration system for daily patient repositioning in proton-therapy. IROS 07, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007.
- [PMW⁺09] N. Padoy, D. Mateus, D. Weinland, M.O. Berger, and N. Navab. Workflow monitoring based on 3D motion features. pages 585–592, 2009.
- [Pop07] Ronald Poppe. Vision-based human motion analysis : An overview. *Computer Vision and Image Understanding*, 108(1-2) :4–18, Octobre 2007.
- [Pow64] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2) :155–162, Janvier 1964.
- [PRD09] Qi Pan, Gerhard Reitmayr, and Tom W. Drummond. Interactive model reconstruction with user guidance. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR '09*, pages 209–210, Washington, DC, USA, 2009. IEEE Computer Society.
- [PS10] Julien Pilet and Hideo Saito. Virtually augmenting hundreds of real pictures : An approach based on learning, retrieval, and tracking. In *VR*, pages 71–78, 2010.
- [PSCF05] Jordi Pagès, Joaquim Salvi, Christophe Collewet, and Josep Forest. Optimised De Bruijn patterns for one-shot shape acquisition. *Image Vision Comput.*, 23 :707–720, Août 2005.
- [PSM09] Miguel Portela Sotelo and Jean-Michel Moreau. Développement d'un module de suivi de mouvements et localisation d'un patient par rapport à un ensemble d'équipements dans une salle de traitement par radiothérapie. Technical Report RR-LIRIS-2009-035, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon, Octobre 2009.
- [PST⁺00] Delphine Plattard, Marine Soret, Jocelyne Troccaz, Patrick Vassal, Jean-Yves Giraud, Guillaume Champleboux, Xavier Artignan, and Michel Bolla. Patient set-up using portal images : 2D/2D image registration using mutual information. *Computer Aided Surgery*, 5(4) :246–262, 2000.
- [PTVF92] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C : the art of scientific computing*. Cambridge University Press, New York, NY, USA, 2nd edition edition, 1992.

- [PYSaVH⁺96] Maxine Washington Paul Y. Song and, Florin Vaida, Russell Hamilton, Danny Spelbring, Brenda Wyman, Joanne Harrison, George T. Y. Chen, and Srinivasan Vijayakumar. A comparison of four patient immobilization devices in the treatment of prostate cancer patients with three dimensional conformal radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 34(1) :213–219, Janvier 1996.
- [RA01] STUK : Säteilyturvakeskus Finnish Radiation and Nuclear Safety Authority. Radiation safety of radiotherapy equipment and treatment rooms. Technical report, STUK, Finland, Février 2001.
- [RBT10] Rim Romdhane, Francois Bremond, and Monique Thonnat. A Framework dealing with Uncertainty for Complex Event Recognition. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS*, Août 2010.
- [Rea97] James Reason. *Managing the risks of organisational accidents*. Ashgate, 1997.
- [Rei79] D. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6) :843–854, Décembre 1979.
- [RK08] A. Joy Rivera and Ben-Tzion Karsh. Human Factors and Systems Engineering Approach to Patient Safety for Radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 71(1, Supplement 1) :174–177, Mai 2008.
- [RL02] Maurice Ringer and Joan Lasenby. Multiple Hypothesis Tracking for Automatic Optical Motion Capture. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 524–536. Springer Berlin / Heidelberg, 2002.
- [Ros02] J. C. Rosenwald. Sécurité en radiothérapie : le contrôle des logiciels et des systèmes informatiques - Safety in radiation therapy : quality assurance of computerized system. *Cancer/Radiothérapie*, 6(Supplement 1) :180–189, 2002.
- [SAN⁺04] Luc Soler, Nicholas Ayache, Stéphane Nicolau, Xavier Pennec, Clément Forest, Hervé Delingette, and Jacques Marescaux. Traitement d’images médicales pour la planification, la simulation et l’aide intra-opératoire des actes chirurgicaux. *La Revue de l’Electricité et de l’Electronique*, pages 64–71, Janvier 2004.
- [SBN⁺09] Jesmin Shafiq, Michael Barton, Douglas Noble, Claire Lemer, and Liam J Donaldson. An international review of patient safety measures in radiotherapy practice. *Radiotherapy and Oncology*, 92(1) :15–21, 2009.
- [SBR⁺06] Maria Francesca Spadea, Guido Baroni, Marco Riboldi, Barbara Tagaste, Cristina Garibaldi, Roberto Orecchia, and Antonio Pedotti. Patient set-up verification by infrared optical localization and body surface sensing in breast radiation therapy. *Radiotherapy and Oncology*, 79(2) :170–178, Mai 2006.

- [SBSR05] Andrew J. Schaefer, Matthew D. Bailey, Steven M. Shechter, and Mark S. Roberts. Modeling Medical Treatment Using Markov Decision Processes. In Margaret L. Brandeau, François Sainfort, and William P. Pierskalla, editors, *Operations Research and Health Care*, volume 70 of *International Series in Operations Research & Management Science*, pages 593–612. Springer US, 2005.
- [SCB⁺02] Christopher F. Serago, Suzanne J. Chungbin, Steven J. Buskirk, Gary A. Ezzell, A. Craig Collie, and Sujay A. Vora. Initial experience with ultrasound localization for positioning prostate cancer patients for external beam radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*, 53(5) :1130–1138, Août 2002.
- [SDV⁺10] Jacques Saadé, Anne-Laure Didier, Pierre-Frédéric Villard, Romain Buttin, Jean-Michel Moreau, Michaël Beuve, and Behzad Shariat. A preliminary study for a biomechanical model of the respiratory system. In *Engineering and Computational Sciences for Medical Imaging in Oncology - VISAPP 2010*, pages 509–515, Mai 2010.
- [SFN08] Tobias Sielhorst, Marco Feuerstein, and Nassir Navab. Advanced Medical Displays : A Literature Review of Augmented Reality. *IEEE/OSA Journal of Display Technology ; Special Issue on Medical Displays*, 4(4) :451–467, Décembre 2008.
- [SFR⁺11] J. Schaerer, A. Fassi, M. Riboldi, G. Baroni, and D. Sarrut1. Time-consistent parametrization form dynamic optical surface detection for respiratory motion recovery in radiation oncology. In *Acts of the 30th European Society for Therapeutic Radiology and Oncology (ESTRO)*, 2011.
- [SG99] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. volume 2, pages 246–252, Los Alamitos, CA, USA, Août 1999. IEEE Computer Society.
- [Sim10] Gilles Simon. In-Situ 3D Sketching Using a Video Camera as an Interaction and Tracking Device. In *31st Annual Conference of the European Association for Computer Graphics - Eurographics 2010*, Norrköping, Suède, Mai 2010.
- [SLB05] Oleksandr A. Skydan, Michael J. Lalor, and David R. Burton. Using coloured structured light in 3-D surface measurement. *Optics and Lasers in Engineering*, 43(7) :801–814, 2005.
- [SMB07] Gaël Sourimant, Luce Morin, and Kadi Bouatouch. GPS, GIS and Video registration for building reconstruction. In *ICIP*, San Antonio, US, September 2007.
- [SMH⁺03] N Smith, I Meir, G Hale, R Howe, L Johnson, P Edwards, D Hawkes, M Bidmead, and D Landau. A comparison of four patient immobilization devices in the treatment of prostate cancer patients with three dimensional conformal radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 57(2) :187, Octobre 2003.
- [Soc10] American Cancer Society. Understanding Radiation Therapy : A Guide for Patients and Families. Technical Report 1, USA, Mars 2010.

- [SPB04] Joaquim Salvi, Jordi Pagès, and Joan Batlle. Pattern Codification Strategies in Structured Light Systems. *PATTERN RECOGNITION*, 37 :827–849, 2004.
- [SRB⁺07] Christian Schmaltz, Bodo Rosenhahn, Thomas Brox, Joachim Weickert, Daniel Cremers, Lennart Wietzke, and Gerald Sommer. Occlusion modeling by tracking multiple objects. In *Proceedings of the 29th DAGM conference on Pattern recognition*, pages 173–183, Berlin, Heidelberg, 2007. Springer-Verlag.
- [SS03] D. S. Shrestha and B. L. Steward. Automatic corn plant population measurement using machine vision. *International Journal of Radiation Oncology*Biology*Physics*, 46(2) :559–565, 2003.
- [SSH10] Jeffrey Schlosser, Kenneth Salisbury, and Dimitre Hristov. Telerobotic system concept for real-time soft-tissue imaging during radiotherapy beam delivery. *Med Phys*, 37(12) :6357–67, 2010.
- [SSW⁺10] SB Scarboro, M Stovall, A White, SA Smith, D Yaldo, SF Kry, and RM Howell. Effect of organ size and position on out-of-field dose distributions during radiation therapy. *Physics in Medicine and Biology*, 53(23) :7025–7036, Décembre 2010.
- [ST94] Jianbo Shi and Carlo Tomasi. Good Features to Track. In *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593–600, 1994.
- [ST01] Cristian Sminchisescu and Bill Triggs. Covariance scaled sampling for monocular 3D body tracking. In *CVPR01*, pages I :447–454, 2001.
- [Sut65] Ivan E. Sutherland. The Ultimate Display. In *Proceedings of IFIPS Congress*, volume 2, pages 506–508, New York, 1965. ACM.
- [Sut68] Ivan E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA, 1968. ACM.
- [TBGG⁺07] BS Teh, C Bloch, M Galli-Guevara, S Richardson L Doh and, S Chiang, P Yeh, M Gonzalez, W Lunn, R Marco, J Jac, AC Paulino, HH Lu, EB Butler, and RJ Amato. The treatment of primary and metastatic renal cell carcinoma (RCC) with image-guided stereotactic body radiation therapy (SBRT). *Biomedical imaging and intervention journal*, 3(1), 2007.
- [TDG⁺00] Frank Tendick, Michael Downes, Tolga Goktekin, Murat Cenk Cavusoglu, David Feygin, Xunlei Wu, Roy Eyal, Mary Hegarty, and Lawrence W. Way. A Virtual Environment Testbed for Training Laparoscopic Surgical Skills. *Presence*, 9(3) :236–255, 2000.
- [TMWG09] J. Talbot, J. Meyer, R. Watts, and R. Grasset. A method for patient set-up guidance in radiotherapy using augmented reality. *Australasian Physical and Engineering Sciences in Medicine*, 32(4), 2009.
- [TSK00] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. A Sampling Algorithm for Tracking Multiple Objects. In *Proceedings of the In-*

- ternational Workshop on Vision Algorithms : Theory and Practice*, ICCV '99, pages 53–68, London, UK, 2000. Springer-Verlag.
- [TWW⁺06] S. Tao, A. Wu, Y. Wu, Y. Chen, and J. Zhang. Patient Set-up in Radiotherapy with Video-based Positioning System. *Clinical Oncology*, 18(4) :363–366, 2006.
- [Uni06] Health Technology Assessment Unit. A Reference Guide for Learning from Incidents in Radiation Treatment. Technical report, Alberta Heritage Foundation for Medical Research (Canada), Canada, Janvier 2006.
- [VBT03] Van-Thinh Vu, François Brémond, and Monique Thonnat. Automatic Video Interpretation : A Recognition Algorithm for Temporal Scenarios Based on Pre-compiled Scenario Models. In *ICVS'03*, pages 523–533, 2003.
- [vdHDT⁺07] Anton van den Hengel, Anthony Dick, Thorsten Thormählen, Ben Ward, and Philip H. S. Torr. VideoTrace : rapid interactive scene modelling from video. *ACM Trans. Graph.*, 26, Juillet 2007.
- [vdHHWD09] Anton van den Hengel, Rhys Hill, Ben Ward, and Anthony Dick. In situ image-based modeling. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '09, pages 107–110, Washington, DC, USA, 2009. IEEE Computer Society.
- [VGK⁺10] Athanasios Voulodimos, Helmut Grabner, Dimitrios Kosmopoulos, Luc Van Gool, and Theodora Varvarigou. Robust workflow recognition using holistic features and outlier-tolerant fused hidden Markov models. In *Proceedings of the 20th international conference on Artificial neural networks : Part I*, ICANN'10, pages 551–560, Berlin, Heidelberg, 2010. Springer-Verlag.
- [VGMVG11] Galina Veres, Helmut Grabner, Lee Middleton, and Luc Van Gool. Automatic workflow monitoring in industrial environments. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part I*, ACCV'10, pages 200–213, Berlin, Heidelberg, 2011. Springer-Verlag.
- [VKK02] J. Velten, M. Krips, and A. Kummert. DSP and FPGA based Implementation of a Video based Industrial Safety Application. In *WSEAS International Conference on Electronics, Control and Signal Processing and E-Activities*, ICECS, Singapore, Décembre 2002.
- [vKP10] D. W. F. van Krevelen and R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality*, 9(2) :1–20, Juin 2010.
- [VLF04] Luca Vacchetti, Vincent Lepetit, and Pascal Fua. Stable Real-Time 3D Tracking Using Online and Offline Information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 :1385–1391, Octobre 2004.
- [VPL⁺97] Eric Vigneault, Jean Pouliot, Jacques Laverdière, Jean Roy, and Marc Dorion. Electronic portal imaging device detection of radioopaque markers for the evaluation of prostate position during megavol-

- tage irradiation : A clinical study. *International Journal of Radiation Oncology*Biography*Physics*, 37(1) :205–212, 1997.
- [VTB⁺06] Daniel Vetterli, Sandrine Thalmann, Frank Behrensmeier, Ludger Kemmerling, Ernst J. Born, Roberto Mini, Richard H. Greiner, and Daniel M. Aebbersold. Daily organ tracking in intensity-modulated radiotherapy of prostate cancer using an electronic portal imaging device with a dose saving acquisition mode. *Radiotherapy and Oncology*, 79(1) :101–108, 2006.
- [VVDM11] Arnaud Verdant, Patrick Villard, Antoine Dupret, and Hervé Mathias. Three novell analog-domain algorithms for motion detection in video surveillance. *J. Image Video Process.*, 2011 :10 :1–10 :13, Janvier 2011.
- [WDM01] Sanghyuk Woo, Albert Dipanda, and Franck Marzani. Application of Genetic Algorithms to 3-D Shape Reconstruction in an Active Stereo Vision System. In *Proceedings of the Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EMCCVPR '01, pages 480–496, London, UK, 2001. Springer-Verlag.
- [Web05] W.A. Weber. Use of PET for monitoring cancer therapy and for predicting outcome. *Journal of Nuclear Medicine*, 46(6) :983–95, 2005.
- [Web06] Steve Webb. Motion effects in (intensity modulated) radiation therapy : a review. *Physics in Medicine and Biology*, 51(13) :R403—R425, Juillet 2006.
- [WKAS78] Sholom M. Weiss, Casimir A. Kulikowski, Saul Amarel, and Aran Safir. A model-based method for computer-aided medical decision-making. *Artificial Intelligence*, 11(1-2) :145–172, 1978.
- [WKP⁺06] Twyla R. Willoughby, Patrick A. Kupelian, Jean Pouliot, Katsuto Shinohara, Michelle Aubin, Mack Roach III, Lisa L. Skrumeda, James M. Balter, Dale W. Litzenberg, Scott W. Hadley, John T. Wei, and Howard M. Sandler. Target localization and real-time tracking using the Calypso 4D localization system in patients with localized prostate cancer. *International Journal of Radiation Oncology*Biography*Physics*, 65(2) :528–534, 2006.
- [WSB09] Daniel Wagner, Dieter Schmalstieg, and Horst Bischof. Multiple target detection and tracking with guaranteed framerates on mobile phones. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '09, pages 57–64, Washington, DC, USA, 2009. IEEE Computer Society.
- [XLD⁺00] L Xing, Z Lin, SS Donaldson, QT Le, D Tate, DR Goffinet, S Wolden, L Ma, and AL Boyer. Dosimetric effects of patient displacement and collimator and gantry angle misalignment on intensity modulated radiation therapy. *Radiotherapy and Oncology*, 56(1) :97–108, Juillet 2000.

-
- [YB03] Milena Yankova and Svetla Boytcheva. Focusing on scenario recognition in information extraction. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 2*, EACL '03, pages 41–48, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.
- [YBC⁺05] Tai Keung Yeung, Karen Bortolotto, Scott Cosby, Margaret Hoar, and Ernst Lederer. Quality assurance in radiotherapy : evaluation of errors and incidents recorded over a 10 year period. *Radiotherapy and Oncology*, 74(3) :283–291, Mars 2005.
- [YLPL05] Tao Yang, Stan Z. Li, Quan Pan, and Jing Li. Real-Time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 970–975, Washington, DC, USA, 2005. IEEE Computer Society.
- [YOI92] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden Markov model. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pages 379–385, 1992.
- [YSB02] Yuling Yan, Yulin Song, and Arthut L. Boyer. An investigation of a video-based patient repositioning technique. *International Journal of Radiation Oncology Biology Physics*, 54(2) :606–614, 2002.
- [Zad65] Lotfi A. Zadeh. Fuzzy Sets. *Information and Control*, 8(3) :338–353, 1965.
- [ZF03] B. Zitova and J. Flusser. Image Registration Methods : A Survey. *IVC*, 21(11) :977–1000, Octobre 2003.
- [Ziv04] Zoran Zivkovic. Improved Adaptive Gaussian Mixture Model for Background Subtraction. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2 - Volume 02*, ICPR '04, pages 28–31, Washington, DC, USA, 2004. IEEE Computer Society.
- [ZN04] Tao Zhao and Ram Nevatia. Tracking Multiple Humans in Complex Situations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 :1208–1221, Septembre 2004.

Références web

- [1] CIRC : Centre international de recherche sur le cancer – IARC : International Agency for Research on Cancer. <http://www.iarc.fr/>.
- [2] ASTRO : American Society for Radiation Oncology. <http://www.astro.org/>.
- [3] ESTRO : European Society for Radiotherapy & Oncology. <http://www.estro.org/>.
- [4] ASN : Autorité de Sûreté Nucléaire. <http://www.asn.fr/>.
- [5] OMS : Organisation mondiale de la santé – WHO : World Health Organization. <http://www.who.int/>.
- [6] SFRO : Société Française de Radiothérapie Oncologique. <http://www.sfro.org/>.
- [7] Centre Léon Bérard. Le Centre de lutte contre le cancer Léon Bérard est un établissement privé à but non lucratif situé à Lyon dans le Rhône en France. <http://www.lyon.fnclecc.fr/>.
- [8] LIRIS : Laboratoire d'Informatique en Image et Systèmes d'information. <http://liris.cnrs.fr/>.
- [9] DOSIsoft. DOSIsoft est le leader français de Systèmes de Planification de Traitement pour la radiothérapie. <http://www.dosisoft.com/>.
- [10] OSEO. Entreprise publique exerçant trois métiers : l'aide à l'innovation, la garantie des concours bancaires et des investisseurs en fonds propres, le financement en partenariat. <http://www.oseo.fr/>.
- [11] Varian. Fournisseur d'équipements et des logiciels pour la radiothérapie. <http://www.varian.com/us/oncology/>.
- [12] Xoft. Fournisseur d'équipements électroniques pour la curithérapie. <http://www.xoftinc.com/>.
- [13] Philips Healthcare. L'activité Healthcare regroupe l'ensemble des systèmes médicaux produits par l'entreprise Philips. <http://www.healthcare.philips.com/>.
- [14] Elekta. Elekta is a human care company pioneering significant innovations and clinical solutions for treating cancer and brain disorders. <http://www.elekta.com/>.
- [15] Alpha cradle®. Alpha Cradle® est une marque enregistrée de la société Smithers Medical Products pour des systèmes d'immobilisation du patient et de repositionnement. <http://www.alphacradle.com/>.
- [16] Orfit. Entreprise qui fabrique des matériaux thermoplastiques pour des applications médicales. <http://www.orfit.com/>.

- [17] AlignRT. Fournisseur de systèmes optiques pour le positionnement, la surveillance et le suivi de la respiration pour des traitement en radiothérapie. <http://www.visionrt.com/>.
- [18] Ndi. NDI est un fournisseur d'appareils de mesure. <http://www.ndigital.com/medical/>.
- [19] Calypso. Le GPS Calypso pour la technologie Body® permet le suivi de la tumeur lors de l'irradiation. <http://www.calypsomedical.com/>.
- [20] ARToolKit. ARToolKit est une bibliothèque logicielle pour le développement des application de réalité augmentée. <http://artoolkit.sourceforge.net/>.
- [21] Steve Kaelher. Fuzzy Logic Tutorial, 1993. <http://seattlerobotics.org/Encoder/mar98/fuz/flindex.html>.
- [22] OpenCV : Open Source Computer Vision. OpenCV est bibliothèque logicielle pour la vision par ordinateur temps réel. <http://opencv.willowgarage.com/>.
- [23] Coin3D. Coin3D est une bibliothèque logicielle, basée sur la bibliothèque OpenGL, pour la gestion de graphes de scène. <http://www.coin3d.org/>.
- [24] OpenGL : Open Graphics Library. OpenGL est API multiplateforme pour la conception d'applications générant des images 2D et 3D. <http://www.opengl.org/>.
- [25] Xerces-C++. Xerces-C++ est un parseur XML développé en C++. <http://xerces.apache.org/xerces-c/>.
- [26] DCMTK : DICOM ToolKit. DCMTK est une bibliothèque logicielle pour le traitement des données au format DICOM. <http://dicom.offis.de/>.
- [27] QT. QT est un framework orienté objet et développé en C++ qui offre des composants d'interface graphique (widgets), d'accès aux données, de connexions réseaux, de gestion des fils d'exécution, d'analyse XML, etc. <http://qt.nokia.com/products/>.
- [28] Blender. Blender est un logiciel libre et gratuit d'animation, de modélisation et de rendu 3D. <http://www.blender.org/>.

TITRE

Vidéosurveillance d'une salle de traitement en radiothérapie

MOTS-CLEFS

Réalité augmentée et virtuelle, modélisation, vision par ordinateur, suivi de mouvement, logique floue, analyse vidéo, reconnaissance de scénarios, assurance qualité, gestion de risques, interactivité, radiothérapie.

TITLE

Video surveillance of a radiotherapy treatment room

KEYWORDS

Augmented and virtual reality, modeling, computer vision, motion tracking, fuzzy logic, video analysis, scenario recognition, quality assurance, risk management, interactivity, radiotherapy.

TÍTULO

Vídeovigilancia de una sala de tratamientos de radioterapia

PALABRAS CLAVES

La realidad aumentada y virtual, modelado digital, la visión artificial, seguimiento del movimiento, lógica difusa, análisis de vídeo, reconocimiento de escenario, aseguramiento de la calidad, gestión de riesgos, interactividad, radioterapia.

ADRESSE

Laboratoire d'InfoRmatique en Image et Systèmes d'information
LIRIS - UMR 5205
Université Claude Bernard - Lyon 1 - DOUA
Bâtiment Nautibus
23-25 Avenue Pierre de Coubertin
69100 Villeurbanne