



**HAL**  
open science

# Approche pixel de la soustraction d'arrière-plan en vidéo, basée sur un mélange de gaussiennes imprécises

Ali Darwich

► **To cite this version:**

Ali Darwich. Approche pixel de la soustraction d'arrière-plan en vidéo, basée sur un mélange de gaussiennes imprécises. Traitement du signal et de l'image [eess.SP]. Université du Littoral Côte d'Opale; École doctorale des Sciences et de Technologie (Beyrouth), 2018. Français. NNT : 2018DUNK0479 . tel-02001886

**HAL Id: tel-02001886**

**<https://theses.hal.science/tel-02001886>**

Submitted on 31 Jan 2019

**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.

École doctorale n° 072 : Sciences Pour l'Ingénieur

# Doctorat Université Lille Nord-de-France

## THÈSE

pour obtenir le grade de docteur délivré par  
**l'Université du Littoral Côte d'Opale**

**Spécialité doctorale : Automatique, Génie informatique,  
Traitement du Signal et des Image**

*présentée et soutenue publiquement par*

**Ali DARWICH**

le 01 Mars 2018

## **Approche pixel de la soustraction d'arrière-plan en vidéo, basée sur un mélange de gaussiennes imprécises**

Directeur de thèse : **André BIGAND**

Directeur de thèse : **Yasser MOHANNA**

### **Jury**

<b>M. Y.MOHANNA,</b>	univ. Libanaise, Professeur	Directeur
<b>M. A.BIGAND,</b>	univ.Littoral (ULCO), MCF(HdR)	Directeur
<b>Mme. S.GALICHET,</b>	univ. Savoie, Professeur	Rapporteur
<b>M. T.BOUWMANS,</b>	univ. La Rochelle, MCF(HdR)	Rapporteur
<b>Mme. J.BENOIS-PINEAU,</b>	univ. de Bordeaux, Professeur	Examinateur, Président jury
<b>M. A.A.HAJ-ALI,</b>	univ. LIU, Professeur	Examinateur
<b>M. O.BAZZI,</b>	univ. Libanaise, Professeur	Examinateur
<b>M. P-A.Hébert,</b>	univ. Littoral (ULCO), MCF	Examinateur

**ULCO**

**Laboratoire d'Informatique Signal et Image de la Côte d'Opale**

Équipe d'Accueil 4491



# Table des matières

<b>Table des matières</b>	<b>iii</b>
<b>Liste des figures</b>	<b>v</b>
<b>Liste des tableaux</b>	<b>vii</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 État de l'art sur la soustraction du fond en vidéo</b>	<b>5</b>
1.1 Introduction . . . . .	7
1.2 Les défis de la soustraction de fond en vidéo . . . . .	8
1.3 Présentation générale . . . . .	10
1.4 Méthodes non itératives . . . . .	10
1.5 Méthodes non itératives de type <i>ACP</i> . . . . .	11
1.6 Méthodes itératives . . . . .	17
1.7 Conclusion du chapitre . . . . .	23
1.8 Références . . . . .	24
<b>2 État de l'art sur l'application des ensembles flous à la soustraction du fond en vidéo</b>	<b>29</b>
2.1 Introduction . . . . .	31
2.2 Incertitudes et imperfections en vidéo . . . . .	31
2.3 Modélisation de l'incertitude par les ensembles flous . . . . .	33
2.4 Méthodes de soustraction de fond basées sur le flou . . . . .	44
2.5 Conclusion du chapitre . . . . .	55
2.6 Références . . . . .	55

<b>3</b>	<b>La méthode <i>IV-FGMM</i> : soustraction du fond basée sur un mélange de gaussiennes imprécises</b>	<b>61</b>
3.1	Introduction . . . . .	63
3.2	Algorithme proposé . . . . .	63
3.3	Décision floue . . . . .	65
3.4	Décisions complémentaires . . . . .	68
3.5	Fusion des décisions . . . . .	70
3.6	Autres étapes . . . . .	71
3.7	Conclusion du chapitre . . . . .	72
3.8	Références . . . . .	72
<b>4</b>	<b>Validation de la méthode</b>	<b>75</b>
4.1	Introduction . . . . .	76
4.2	Indicateurs de performance . . . . .	76
4.3	Validation des étapes de la méthode . . . . .	77
4.4	Comparaison à des méthodes basées sur le <i>GMM</i> et le flou . . . . .	82
4.5	Comparaison à des méthodes basées sur l' <i>ACP</i> . . . . .	86
4.6	Comparaison des temps de calcul des différentes méthodes . . . . .	87
4.7	Conclusion du chapitre . . . . .	89
4.8	Références . . . . .	90
	<b>Conclusion générale et perspectives</b>	<b>93</b>
<b>A</b>	<b>Annexe</b>	<b>I</b>
A.1	Tableaux de résultats pour toutes les méthodes basées sur le <i>GMM</i> . . . . .	I
A.2	Tableaux de résultats pour toutes les méthodes basées sur l' <i>ACP</i> . . . . .	IV

# Liste des figures

1.1	<b>Diagramme d'un algorithme générique de soustraction d'arrière-plan(ou fond).</b> . . . . .	7
1.2	<i>ACP</i> classique sans point aberrant. . . . .	12
1.3	<i>ACP</i> classique avec points aberrants. . . . .	13
1.4	<i>ACP</i> robuste sans point aberrant. . . . .	15
1.5	<i>ACP</i> robuste avec points aberrants. . . . .	15
1.6	ALM (a) <i>Fond</i> (d) <i>Objet</i> , EALM (b) (e), FPCP (c) (f), IALM (m) (n), RPCA (g) (j), LSADM (h) (k), MoG-ACP (i) (l). . . . .	16
1.7	<b>Algorithme de la méthode de mélange de gaussiennes.</b> . . . .	19
2.1	Interprétations de l'incertitude dans l'espace « FIR ». . . . .	33
2.2	Ensemble flou vs ensemble net (intervalle) . . . . .	34
2.3	Nombres flous et coupe de niveau $\alpha$ . . . . .	35
2.4	Ensembles flous intuitifs avec fonction d'appartenance et de non-appartenance. . . . .	37
2.5	Ensemble flou intuitif d'Atanassov $\mu_A(x) + \nu_A(x) \leq 1 \forall x \in X$ . . . . .	38
2.6	Les espaces d'incertitude des <i>PFN</i> et <i>IFN</i> . . . . .	39
2.7	Fonction d'appartenance secondaire d'un ensemble flou de type-2 pour $x = x'$ . . . . .	41
2.8	<i>IVFS</i> d'un mode gaussien (incertitude sur la moyenne). . . . .	47
2.9	<i>IVFS</i> d'un mode gaussien (incertitude sur l'écart-type). . . . .	48
2.10	Insertion du flou dans un mode gaussien. . . . .	49
2.11	<b>Algorithme de la méthode T2-FGMM.</b> . . . .	49
2.12	Images 1 et 100 d'une séquence vidéo, avec marquage du pixel étudié. . . . .	50
2.13	Évolution de l'intensité du pixel sur les 100 premières images (en noir) et état du <i>GMM</i> à l'image 100 (en rouge). . . . .	50
2.14	Deux images de la vidéo <i>UL</i> , avant et après le changement d'éclairage. . . . .	52

2.15	<i>GMM</i> net à l'image 1, relatif au pixel marqué sur l'image précédente. . . . .	53
2.16	<i>GMM IVFS</i> à l'image 1, relatif au pixel étudié. . . . .	53
2.17	<i>GMM</i> net à l'image 500, relatif au pixel étudié. . . . .	54
2.18	<i>GMM IVFS</i> à l'image 500, relatif au pixel étudié. . . . .	54
3.1	<b>Algorithme de la méthode proposée.</b> . . . . .	64
3.2	Intervalles flous des modes gaussiens, avant agrégation. . . . .	66
3.3	Intervalles flous agrégés + Zone de rejet. . . . .	66
3.4	Exemples de décisions floues par comparaison des intervalles agrégés. . . . .	68
3.5	Fusion spatiale : poids des décisions floues initiales. . . . .	69
3.6	Fusion spatiale : agrégation des décisions floues initiales. . . . .	69
4.1	Image initiale (a), Binaire (0.564) (b), Binaire/temporelle (0.685) (c), Binaire/spatiale(0.915)(d), Binaire/spatio-temporelle(0.911)(e), Vérité terrain(f), Floue(0.574)(g), Floue/temporelle(0.667)(h), Floue/spatiale(0.9)(i), Floue/spatio-temporelle(0.943), (j). . . . .	80
4.2	Spécificité selon les variantes, vidéo Canoé. . . . .	80
4.3	PCB selon les variantes, vidéo Canoé. . . . .	81
4.4	Rappel selon les variantes, vidéo Canoé. . . . .	81
4.5	Précision selon les variantes, vidéo Canoé. . . . .	82
4.6	F-mesures selon les variantes, vidéo Canoé. . . . .	82
4.7	Temps de calcul en fonction du nombre d'images de la séquence vidéo. . . . .	89

# Liste des tableaux

4.1 Classement des décisions. . . . .	76
4.2 Indicateurs de performance. . . . .	77
4.3 Paramètres de la méthode . . . . .	78
4.4 Jeu de paramètres optimisé pour la vidéo canoé . . . . .	78
4.5 Performances de chaque variante, vidéo Canoé. . . . .	79
4.6 Jeu de paramètres optimisé sur la catégorie de vidéos <i>fond dynamique</i> . . . . .	83
4.7 Méthodes utilisées dans l'étude comparative. . . . .	84
4.8 Paramètres optimisés pour les méthodes <i>T2-FGMM</i> . . . . .	85
4.9 F-mesures de l'ensemble des vidéos. . . . .	85
4.10 Paramètres optimisés sur les vidéos réduites. . . . .	86
4.11 F-mesures des méthodes de type ACP. . . . .	87
4.12 Temps de calcul des différents méthodes ( $t_{lim} = 20\ 000s$ ). . . . .	88
A.1 Vidéo boats . . . . .	I
A.2 Vidéo Canoé. . . . .	II
A.3 Vidéo Fall. . . . .	II
A.4 Vidéo Overpass. . . . .	II
A.5 Vidéo Fontaine1. . . . .	III
A.6 Vidéo Fontaine2. . . . .	III
A.7 Vidéo Canoé . . . . .	IV
A.8 Vidéo Overpass . . . . .	IV
A.9 Vidéo Fontaine1 . . . . .	IV
A.10 Vidéo Fontaine2 . . . . .	V
A.11 Vidéo boats . . . . .	V
A.12 Vidéo fall. . . . .	V



## **Liste des publications liées à la thèse**

### **Communications dans des Congrès Internationaux avec Comité de Lecture**

- Ali Darwich, Pierre-Alexandre Hébert, André Bigand, Yasser Mohanna. “Background Subtraction under Uncertainty using a Type-2 Fuzzy Set Gaussian Mixture Model”. Congrès IEEE CSEET’17, Liban, 2017.
- Ali Darwich, Pierre-Alexandre Hébert, André Bigand, Yasser Mohanna. “Analyse en composantes principales, application a la modélisation de l’arrière-plan”, JFL3 Liban, 2016.

## Résumé

La détection d'objets en mouvement représente une étape très importante pour de nombreuses applications telles que l'analyse du comportement humain pour la surveillance visuelle, la reconnaissance d'action par modèle, la suivi du trafic routier, etc. La soustraction d'arrière-plan est une approche populaire, mais difficile étant donnée qu'elle doit surmonter de nombreux obstacles, comme l'évolution dynamique du fond, les variations de luminosité, les occlusions, etc.

Dans les travaux présentés, nous nous sommes intéressés à ce problème de segmentation objets / fond, avec une modélisation floue de type-2 pour gérer l'imprécision du modèle et des données. La méthode proposée modélise l'état de chaque pixel à l'aide d'un modèle de mélange de gaussiennes imprécis et évolutif, qui est exploité par plusieurs classifieurs flous pour finalement estimer la classe du pixel à chaque image. Plus précisément, cette décision prend en compte l'historique de son évolution, mais aussi son voisinage spatial et ses éventuels déplacements dans les images précédentes. Puis nous avons comparé la méthode proposée avec d'autres méthodes proches, notamment des méthodes basées sur un modèle de mélanges gaussiens, des méthodes basées floues, ou de type *ACP*. Cette comparaison nous a permis de situer notre méthode par rapport à l'existant et de proposer quelques perspectives à ce travail.

## Abstract

Moving objects detection is a very important step for many applications such as human behavior analysis for visual surveillance, model-based action recognition, road traffic monitoring, etc. Background subtraction is a popular approach, but difficult given that it must overcome many obstacles, such as dynamic background changes, brightness variations, occlusions, and so on.

In the presented works, we focused on this problem of objects/background segmentation, using a type-2 fuzzy modeling to manage the inaccuracy of the model and the data. The proposed method models the state of each pixel using an imprecise and scalable Gaussian mixture model, which is exploited by several fuzzy classifiers to ultimately estimate the pixel class at each image. More precisely, this decision takes into account the history of its evolution, but also its spatial neighborhood and its possible displacements in the preceding images. Then we compared the proposed method with other close methods, including methods based on a gaussian mixture model, fuzzy based methods, or *ACP* type methods. This comparison allowed us to assess its good performances, and to propose some perspectives to this work.

# Introduction générale

## Contexte

L'urbanisme grandissant et l'augmentation des populations génèrent de nouveaux challenges pour les chercheurs, en particulier dans le domaine du traitement des images. La circulation des véhicules et des personnes dans la plupart des villes est très dense. D'autre part, la sécurité des personnes et des biens est aussi une problématique importante. Dans ce contexte, un réseau de vidéo-surveillance des lieux publics et privés est primordial, autant que des algorithmes capables de traiter automatiquement les vidéos produites. Cette étude s'inscrit pleinement dans ce cadre, mais en restreignant son objectif à un traitement vidéo crucial : la détection non-supervisée des objets, et son corollaire, la reconnaissance de l'arrière-plan.

L'extraction du fond, également connu sous le nom de détection de premier plan, ou soustraction de fond, est une technique classique dans le domaine du traitement de l'image et de la vision par ordinateur, dans lequel le premier plan d'une image (souvent un ensemble d'objets en mouvement) est extrait pour un traitement en-ligne ou hors-ligne (de reconnaissance d'objets, etc.). Généralement les régions d'une image d'intérêt sont des objets (humains, voitures, textes, etc.).

Ce problème peut être résolu grossièrement avec des algorithmes très basiques (basées sur la différence entre un modèle d'image-fond et l'image courante), mais les vidéos contiennent fréquemment des éléments perturbateurs susceptibles de mettre en défaut de telles méthodes. En particulier, les arrière-plans dynamiques, ou les variations de luminosité dans les vidéos, peuvent fortement complexifier la classification d'un pixel en *objet* ou *fond*.

Pour traiter de telles vidéos complexes (souvent en lumière naturelle), des méthodes de soustraction de fond basées sur une modélisation plus évoluée ont par conséquent été développés. Une des premières solutions proposées a consisté à utiliser un modèle de mélange de gaussiennes (Gaussian mixture models « *GMM* » en anglais); ce modèle demeure encore largement répandu, et continue de faire l'objet de nombreuses extensions. Plus récemment, ce problème de soustraction de fond a été traité par une catégorie très intéressante de méthode, les *ACP* robustes, qui se sont révélées très performantes.

Il s'agit donc d'un problème de classification, pour lequel il existe déjà une multitude d'algorithmes. Mais leur grande majorité sont principalement destinés à des données précises. Or les vidéos peuvent être fortement bruitées, et les paramètres des modèles peuvent être compliqués à estimer de manière précise. Il peut donc être intéressant de prendre en compte cette imprécision dans le processus même de classification (niveaux de confiance dans le partitionnement), en proposant des méthodes de classification imprécises ou alternatives. En particulier les avantages des méthodes floues (supervisées ou non-supervisées)

pour le traitement des données imprécises sont connus depuis longtemps.

Au cours des dernières années, de nombreux chercheurs ont exploité la théorie des sous-ensembles flous pour développer de nouveaux algorithmes de traitement d'images. Il y a beaucoup de raisons qui rendent cette théorie particulièrement adaptée au traitement d'images. La raison la plus importante est qu'elle fournit un cadre mathématique adéquat pour la représentation et le traitement de connaissances d'experts. La deuxième raison est que les incertitudes au sein des tâches de traitement d'image ne sont pas toujours de nature aléatoire pure, mais ont souvent comme origine l'imprécision et l'ambiguïté, que les techniques floues permettent de gérer de manière efficace.

Dans cette étude, nous proposons un algorithme non-supervisé de soustraction de fond, où l'incertitude est modélisée par un mélange de gaussiennes floues de type-2. Ensuite nous proposons d'exploiter la continuité spatio-temporelle, en construisant plusieurs classifieurs, exploitant chacun des informations complémentaires, dans le but de produire une décision floue *Fond/Objet* aussi sûre que possible. La méthode fait finalement l'objet d'une validation, basée sur des vidéos de référence spécifiquement dédiées à cette problématique d'extraction de fond dans des cas complexes.

## Organisation de la thèse

Pour rendre compte de nos travaux, nous commençons dans le chapitre 1 par proposer une étude bibliographique des méthodes classiques traitant de la soustraction du fond, telles que les méthodes basées sur un mélange de gaussiennes, et les méthodes de réduction de dimensions (*ACP* classique et robustes).

Dans le chapitre 2 nous présentons la modélisation de l'incertitude utilisant les ensembles flous, tels que les ensembles flous de type-1, les ensembles flous pythagoriciens, et les ensembles flous de type-2. Nous détaillons ensuite la méthode *T2GMM*, qui combine modèle *GMM* et ensembles flous de type-2.

Dans le chapitre 3, nous détaillons la méthode de soustraction d'arrière-plan que nous proposons, dont l'originalité réside dans sa gestion de l'incertitude, basée sur un mélange de gaussiennes et sur des ensembles flous de type-2.

Enfin le chapitre 4 décrit le protocole et les résultats de l'évaluation de notre méthode, conduite sur la base de données de référence *ChangeDetection2014*, notamment composée de vidéos à fond dynamique.

## Contributions de la thèse

La détection d'objets dans une vidéo numérique est l'un des principaux challenges de nombreuses applications de vidéo-surveillance. Dans cette thèse, nous nous sommes intéressé au problème difficile de la détection non-supervisée d'objets dans des vidéos dont le fond est en mouvement permanent. Dans ce cas il y a un risque élevé de classer un pixel du fond comme un objet, par exemple dans une vidéo comprenant de l'eau en mouvement en arrière-plan.

Notre objectif est de développer une méthode de soustraction du fond capable de gérer le bruit affectant nécessairement l'estimation des paramètres d'un *GMM*. On choisit de gérer l'incertitude à l'aide d'un modèle flou de type-2, dans le but d'éviter le problème du classement aberrant des pixels qui appartiennent

à une partie dynamique du fond.

La méthode proposée fournit une première classification floue, pixel par pixel. Puis elle conforte cette décision en utilisant le voisinage spatio-temporel de chaque pixel, de manière à récolter plus d'informations sur celui-ci. La classification finalement obtenue est donc le résultat d'une fusion de classifications floues, que l'on espère plus robuste aux perturbations du fond.



# Chapitre 1

## État de l'art sur la soustraction du fond en vidéo

### Sommaire

---

<b>1.1 Introduction</b> . . . . .	<b>7</b>
<b>1.2 Les défis de la soustraction de fond en vidéo</b> . . . . .	<b>8</b>
<b>1.3 Présentation générale</b> . . . . .	<b>10</b>
<b>1.4 Méthodes non itératives</b> . . . . .	<b>10</b>
1.4.1 Filtre médian temporel . . . . .	10
1.4.2 Estimation par la méthode de Parzen . . . . .	11
<b>1.5 Méthodes non itératives de type ACP</b> . . . . .	<b>11</b>
1.5.1 ACP classique . . . . .	11
1.5.2 ACP « tensorielle » . . . . .	13
1.5.3 ACP Robuste . . . . .	13
<b>1.6 Méthodes itératives</b> . . . . .	<b>17</b>
1.6.1 Différence entre images (Frame differencing) . . . . .	17
1.6.2 Méthodes basées sur la soustraction d'une « image-fond » . . . . .	18
1.6.3 Méthodes basées sur un mélange de gaussiennes . . . . .	19
1.6.4 BMOG : Modèle de mélange gaussien « boosté » avec une complexité contrôlée ( Boosted Gaussian Mixture Model with Controlled Complexity) . . . . .	22
1.6.5 Mélange gaussien basé-région pour la modélisation dynamique de fond RMOG (Region-based Mixture of Gaussians) . . . . .	22

1.6.6 Méthodes utilisant le flot optique . . . . .	22
<b>1.7 Conclusion du chapitre . . . . .</b>	<b>23</b>
<b>1.8 Références . . . . .</b>	<b>24</b>

---

## 1.1 Introduction

Dans de nombreuses applications de vision par ordinateur, l'identification d'objets en mouvement est une tâche essentielle. Une partie importante de cette tâche consiste en la classification des pixels des images de la vidéo en premier plan (objets mobiles) ou arrière-plan (fond fixe). C'est souvent un préalable à des applications de vision diverses, comme la reconnaissance de visages, le suivi de véhicules, la télédétection, la vidéo-surveillance ou la surveillance du trafic routier.

Il existe de nombreux algorithmes d'extraction de fond, la plupart d'entre-eux suivant un diagramme de flux simple, composé de quatre étapes majeures : pré-traitement, modélisation d'arrière-plan, détection de premier plan et validation de données. La détection et le suivi des objets en mouvement peuvent être considérés comme des tâches de vision de niveau inférieur, destinées à préparer la compréhension de l'événement de niveau supérieur (prise de décision).

L'approche commune consiste à effectuer une soustraction d'arrière-plan, qui identifie les objets en mouvement à partir d'un masque vidéo qui varie de façon significative (modèle d'arrière-plan). Il existe de **nombreux défis** dans le développement d'un bon algorithme de soustraction d'arrière-plan. Tout d'abord, il doit être robuste vis à vis des changements d'éclairage. Deuxièmement, il doit éviter de détecter des objets d'arrière-plan non stationnaires tels que des feuilles oscillantes, de la pluie, de la neige et des ombres provenant d'objets en mouvement. Enfin, le modèle d'arrière-plan interne doit réagir rapidement aux changements de fond tels que le démarrage et l'arrêt des véhicules par exemple. Un bon algorithme de soustraction d'arrière-plan doit gérer les objets en mouvement qui se fondent d'abord en arrière-plan, puis apparaissent au premier plan plus tard. De nouveaux algorithmes viennent régulièrement s'ajouter à la littérature, le problème de l'identification des objets en mouvement dans un environnement complexe étant encore loin d'être complètement résolu.

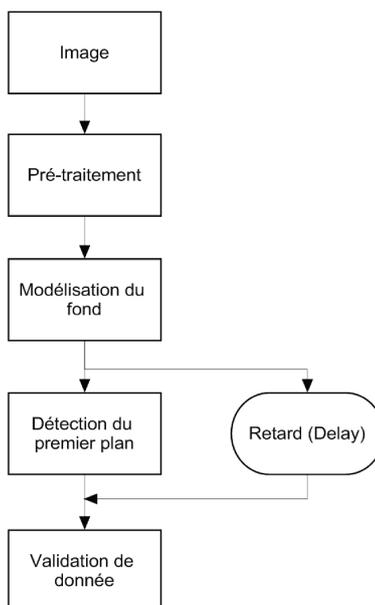


FIGURE 1.1 – Diagramme d'un algorithme générique de soustraction d'arrière-plan (ou *fond*).

La plupart des algorithmes utilisés suivent le schéma simple illustré à la figure 1.1, le pré-traitement consiste en une collection de tâches de traitement d'image simples, qui transforment la vidéo d'entrée

brute en un format plus à même d'être traité par les étapes ultérieures, de plus haut niveau. La modélisation d'arrière-plan utilise la nouvelle image vidéo pour calculer et mettre à jour un modèle de fond. Ce modèle d'arrière-plan fournit une description statistique de l'ensemble de la scène de fond. La détection du premier plan identifie les pixels dans le cadre vidéo qui ne peuvent pas être expliqués de manière adéquate par le modèle d'arrière-plan, et intègrent ces pixels dans un masque de premier plan. Enfin, la validation des données examine le masque candidat, élimine les pixels qui ne correspondent pas aux objets en mouvement réel et produit le masque final de premier plan.

La modélisation d'arrière-plan est le bloc le plus important dans tout algorithme de soustraction de fond. Beaucoup de recherches ont été consacrées à l'élaboration d'un modèle de fond qui soit suffisamment robuste aux changements environnementaux en arrière-plan, mais qui reste sensible pour identifier tous les objets en mouvement dignes d'intérêt. Nous classons les techniques de modélisation de fond en deux grandes catégories, itératives et non itératives.

## 1.2 Les défis de la soustraction de fond en vidéo

La détection et le suivi d'objets restent un problème de recherche ouvert, même après plusieurs années de travail dans ce domaine. Aujourd'hui, le défi en extraction de fond consiste à développer des approches robustes, précises et performantes, sur des vidéos complexes.

Le niveau de difficulté de ce problème dépend fortement de la façon dont on définit l'objet à détecter et à suivre. Si la représentation d'un objet se limite à quelques caractéristiques visuelles (par exemple la couleur), alors il n'est pas si difficile d'identifier et regrouper tous les pixels partageant ces mêmes caractéristiques. Cependant, il y a toujours un risque de confusion avec un autre objet ou élément d'arrière-plan ayant les mêmes attributs. De plus, le changement d'éclairage dans la scène ne garantit pas que ces caractéristiques (la couleur par exemple) se maintiendront pour le même objet dans toutes les trames. Cette segmentation sur quelques caractéristiques visuelles seules conduit à des résultats approximatifs.

D'après [BOUWMANS et ZAHZAH \[2014\]](#), trois conditions principales sont nécessaires au bon fonctionnement de la soustraction de fond en vidéo-surveillance : la caméra doit être fixe, l'éclairage constant et l'arrière-plan statique.

En pratique, diverses perturbations peuvent venir mettre à mal ces conditions, et constituent autant de défis pour la soustraction de fond :

- **Bruit dans l'image** : il est souvent dû à une source d'images de mauvaise qualité (par exemple, images issues de webcam, ou fortement compressées).
- **Gigue de caméra** : la vidéo peut être capturée par des caméras instables (par exemple, à cause de vibrations). L'amplitude de la gigue varie d'une vidéo à l'autre. De même dans certaines conditions, le vent peut provoquer le balancement de la caméra et provoquer ainsi un mouvement nominal dans la séquence.
- **Réglages automatiques de la caméra** : la plupart des appareils photo modernes disposent de la mise au point automatique, du contrôle automatique du gain, de la balance des blancs automatique et du

contrôle automatique de la luminosité. Ces ajustements modifient la dynamique dans les niveaux de couleur entre les différentes images de la séquence.

- **Changements de luminosité** : ils peuvent être relativement progressifs, comme dans une scène en plein air (déplacement du soleil), ou soudains comme l'allumage d'une lampe dans une scène d'intérieur. L'illumination affecte fortement l'apparence du fond et provoque de fausses détections.
- **Bootstrap** : pour les techniques nécessitant un apprentissage, les images d'apprentissage ne décrivent pas toujours l'arrière-plan de manière exhaustive. Il est alors impossible de calculer une image de fond représentative.
- **Fond dynamique** : certaines parties du décor peuvent contenir un mouvement (une fontaine, les mouvements des nuages, le balancement des branches d'arbres, une vague d'eau, etc.), et doivent néanmoins être considérées comme arrière-plan. Un tel mouvement peut être périodique ou irrégulier (par exemple, feux de circulation, arbres ondulants). Gérer une telle dynamique de fond est une tâche très difficile.
- **Camouflage** : intentionnellement ou non, certains objets peuvent s'apparenter au fond environnant. C'est un problème particulièrement délicat dans les applications de surveillance. Et il affecte notamment les méthodes de différenciation temporelle.
- **Ouverture de premier plan** : lorsqu'un objet déplacé contient des régions uniformément colorées, il y a un risque que les pixels de ces régions ne soient pas détectés comme mouvants, et soient catégorisés comme premier plan.
- **Objets d'arrière-plan en mouvement (déplacés)** : l'arrière-plan peut être constitué d'« objets », susceptibles d'être déplacés. Dans l'idéal, ces objets ne devraient pas être considérés comme faisant partie du premier plan.
- **Objets d'arrière-plan insérés** : un nouvel objet d'arrière-plan peut être inséré. Suivant le contexte, ces objets ne doivent pas être considérés comme faisant partie du premier plan.
- **Début d'un objet en mouvement** : lorsqu'un objet initialement confondu avec l'arrière-plan se met en mouvement, les parties de l'arrière-plan alors révélées sont appelées « fantômes » ; elles devraient être rapidement assimilées à l'arrière-plan.
- **Objet de premier plan endormi** : un objet de premier plan qui devient immobile, risque de s'incorporer dans l'arrière-plan. Suivant le contexte, cela n'est pas forcément souhaitable.
- **Ombres** : les ombres peuvent être détectées comme premier plan, et peuvent provenir aussi bien d'éléments du fond ou d'objets en mouvement [AMATO et collab. \[2014\]](#).
- **Météo difficile** : la détection d'un objet en mouvement devient un travail très difficile lorsque les vidéos sont capturées dans des conditions météorologiques défavorables, comme en hiver, avec par exemple des tempêtes de neige, de la neige au sol, du brouillard, des vents violents (turbulences), etc.

## 1.3 Présentation générale

Depuis 1999 [STAUFFER et GRIMSON \[1999\]](#), la littérature propose un très grand nombre de méthodes dans le domaine. Pour faciliter l'analyse et la comparaison de ces méthodes, plusieurs auteurs se sont intéressés à la manière de les catégoriser [BOUWMANS \[2009, 2011, 2014\]](#); [BOUWMANS et ZAHZAH \[2014\]](#); [ELHABIAN et collab. \[2008\]](#); [JEEVA et SIVABALAKRISHNAN \[2015\]](#). On peut donc trouver une grande diversité d'axes de classification : méthodes paramétriques et non-paramétriques, itératives et non-itératives, linéaires et non-linéaires, rapides ou non, etc.

Dans nos travaux nous avons pris le parti d'utiliser une distinction plutôt usuelle, celle qui oppose les approches itératives à celles non itératives. Dans les sections suivantes, nous présentons donc les méthodes les plus classiques, organisées suivant ces 2 catégories.

## 1.4 Méthodes non itératives

L'approche non itérative consiste à estimer le fond à partir d'une collection finie d'images, de manière globale et simultanée.

Cela ne signifie pas pour autant qu'il leur est impossible de s'adapter à un changement de dynamique du fond. Il suffit en effet de renouveler le modèle du fond à chaque nouvelle image, en opérant sur une collection d'images récentes, constituée à l'aide d'une fenêtre coulissante. L'algorithme non-itératif doit alors conserver en mémoire le tampon des  $n$  images précédentes de la vidéo.

Les techniques non itératives sont même hautement adaptables puisqu'elles ne dépendent pas de l'historique au-delà des images stockées dans le tampon. Cependant, l'exigence de stockage peut-être significative lorsqu'un grand tampon est nécessaire pour faire face à des mouvements lents. Ce problème peut être partiellement atténué en stockant les images vidéo à un taux d'image inférieur.

Les techniques non itératives les plus répandues sont décrites ci-dessous.

### 1.4.1 Filtre médian temporel

Dans le filtre médian temporel [LO et VELASTIN \[2001\]](#), le modèle de fond est d'abord construit en calculant la valeur médiane des niveaux de chaque pixel sur les  $n$  dernières images :

$$I_{med}(x_t^p) = med(x_{t-1}^p, x_{t-2}^p, \dots, x_{t-n}^p), \quad (1.1)$$

avec  $x_{t-i}^p$  l'intensité du pixel  $p$  dans la  $i^e$  image précédente, et  $med(.)$  l'opérateur de médiane.

Ensuite, un nouveau pixel est comparé à celui du modèle : il est considéré comme fond si et seulement si la valeur absolue de sa différence tombe sous un certain seuil. La méthode a été améliorée par [Cucchiara et al.](#) dans [CALDERARA et collab. \[2006\]](#) et [CUCCHIARA et collab. \[2001\]](#), en remplaçant le traitement indépendant des composantes couleur (par médiane), par un traitement conjoint (par médoïde).

Le filtre médian temporel offre une précision acceptable tout en autorisant une fréquence d'images élevée, sous des contraintes de mémoire limitée [PICCARDI \[2004\]](#). Mais elle n'est pas non plus capable de gérer les fonds dynamiques.

## 1.4.2 Estimation par la méthode de Parzen

Elgammal *et al.* [ELGAMMAL et collab. \[2000\]](#) ont proposé un modèle qui s'adapte rapidement aux changements de la scène, permettant ainsi une détection très sensible des cibles mobiles.

Contrairement aux méthodes précédentes, le modèle du fond est multimodal, ce qui permet de traiter correctement les fonds dynamiques.

La méthode repose sur une estimation de la distribution des niveaux d'arrière-plan de chaque pixel, par la méthode de Parzen (*KDE*, en anglais). Soit  $x_1^p, \dots, x_{n-1}^p, x_n^p$ , l'historique des  $n$  derniers niveaux du pixel  $p$ , la distribution du fond à l'instant  $t$  est estimée de la manière suivante :

$$f_t(x) = \frac{1}{n} \sum_{k=1}^n K(x - x_k^p), \quad (1.2)$$

avec  $K$  un noyau gaussien.

Un pixel  $p$  est étiqueté comme un premier plan s'il est peu vraisemblable qu'il provienne de cette distribution, c'est-à-dire si  $f_t(x_t^p)$  est inférieur à un seuil prédéfini.

L'utilisation systématique des  $n$  dernières images pour estimer la fonction de densité permet au modèle d'oublier rapidement le passé.

## 1.5 Méthodes non itératives de type ACP

La réduction de dimension consiste en la transformation des données d'origine vers un espace de dimension inférieure, tout en préservant leur structure spatiale. Plusieurs techniques de réduction de dimension existent, dont notamment l'analyse en composantes principales, la poursuite de projection, et l'algorithme des courbes principales.

L'analyse en composantes principales *ACP* (ou *PCA* en anglais, pour *principal component analysis*) a été développée au début du *XIX*<sup>e</sup> siècle pour analyser des données de grands volumes à cette époque. C'est une technique de compression de données, qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables décorrélées les unes des autres. Ces nouvelles variables sont nommées « composantes principales ». L'*ACP* permet au praticien de réduire le nombre de variables en éliminant l'information redondante, pour faciliter une analyse ultérieure.

### 1.5.1 ACP classique

La méthode est basée sur une analyse de la matrice de covariance ou de corrélation [POP \[2001\]](#). Dans l'*ACP* classique, on réduit la dimension d'un nuage de points centrés, en le projetant dans un sous-espace de dimension inférieure.

En rangeant les coordonnées du nuage de points dans une matrice  $X$ , le problème s'écrit alors :

$$X = L + S, \quad (1.3)$$

où  $L$  est une matrice de rang inférieur à  $X$ , obtenue par projection de  $X$ , et  $S$  est une matrice de pertur-

bations, à minimiser. L'ACP cherche la matrice  $L$  de rang  $k$  des coordonnées d'une projection du nuage  $X$ , qui soit aussi proche que possible du nuage initial, au sens de la distance euclidienne :  $L = X.P$ , avec  $P$  une matrice orthonormée.

Ce problème peut être résolu efficacement par la décomposition en valeurs singulières (SVD).

Dans son application à la soustraction du fond, l'idée consiste à récupérer l'arrière-plan dans la matrice  $L$  de rang faible, tandis que les pixels-objets sont les valeurs aberrantes rares capturées par  $S$ . Pour obtenir le masque du premier plan,  $S$  doit être seuillée, le seuil devant être déterminé expérimentalement.

### Application sous MATLAB™

Pour illustrer l'analyse des données en composantes principales, nous avons utilisé le logiciel MATLAB™. Nous avons généré un nuage de points elliptique de manière aléatoire, et nous lui avons appliqué l'algorithme de calcul des composantes principales, puis de reconstitution des données dans l'espace initial. Les résultats sont présentés sur les figures 1.2 et 1.3.

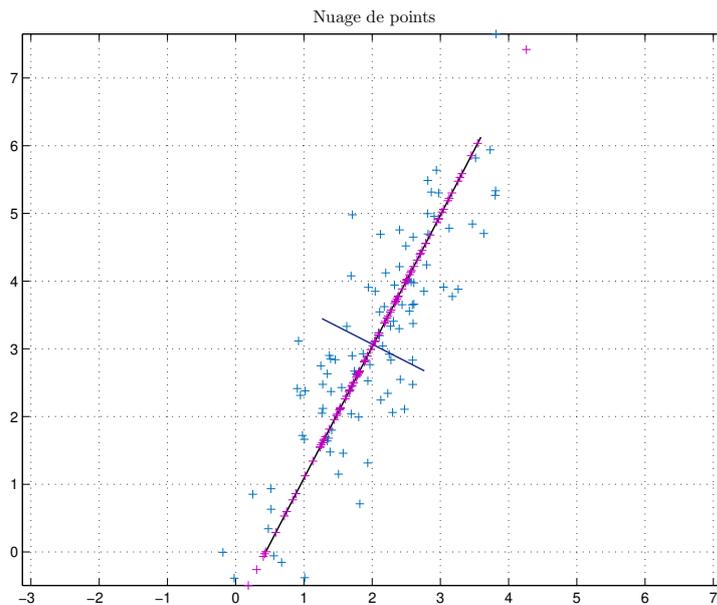


FIGURE 1.2 – ACP classique sans point aberrant.

La figure 1.2 représente le nuage de points, avec ses axes principaux, ainsi que la reconstruction des données après élimination de la plus faible composante (en rouge). On observe que la reconstruction est ici tout à fait satisfaisante.

Après ce premier exemple, nous avons ajouté des points aberrants au nuage de points (cf. figure 1.3). Nous pouvons voir qu'il y a une rotation considérable des axes principaux à cause des valeurs aberrantes ajoutées, ce qui donne une mauvaise reconstruction des données : les projections (en rouge) sont éloignées des points initiaux (en bleu).

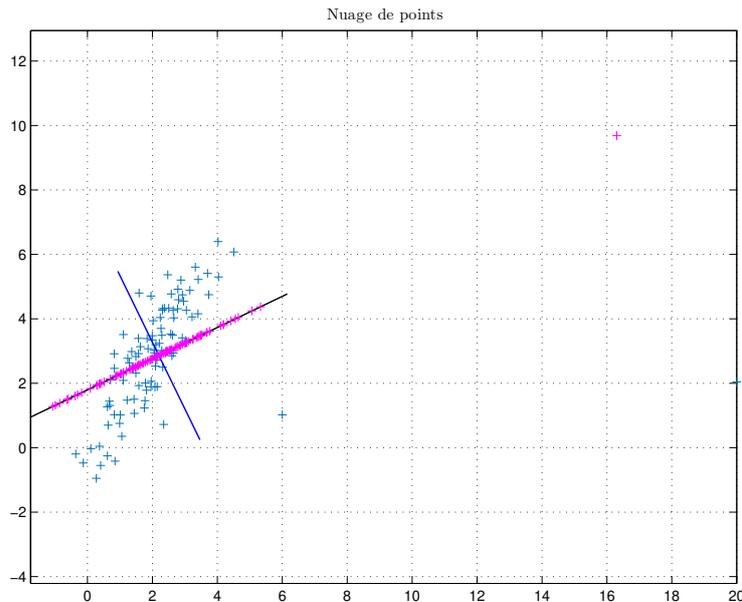


FIGURE 1.3 – ACP classique avec points aberrants.

### 1.5.2 ACP « tensorielle »

L'analyse traditionnelle des données matricielles est intrinsèquement bidimensionnelle, ce qui limite sa capacité à extraire des informations dans une perspective multidimensionnelle. D'autre part, l'analyse de données multilinéaires à base de tenseurs a montré que les modèles tensoriels sont capables de tirer pleinement parti des structures multilinéaires pour fournir une meilleure compréhension et plus de précision. Plusieurs algorithmes importants pour le calcul de l'ACP basés sur la décomposition de tenseurs ont été récemment développés [CHI et KOLDA \[2012\]](#); [LI et collab. \[2010\]](#); [LIU et collab. \[2012\]](#); [SOBRAL et collab. \[2015b\]](#); [ZHOU et collab. \[2012\]](#). Ces modèles semblent promis à un développement important dans le futur.

### 1.5.3 ACP Robuste

L'analyse robuste en composantes principales est une variante largement utilisée de l'analyse en composantes principales classique, qui est moins sensible aux observations manifestement corrompues.

Plusieurs approches d'ACP robustes ont été développées, toutes visant à décomposer la matrice de données en deux matrices, une de faible rang (L), l'autre parcimonieuse (S) [BABACAN et collab. \[2012\]](#); [CANDÈS et collab. \[2011\]](#); [GUYON et collab. \[2012a,b, 2013\]](#); [DE LA TORRE et BLACK \[2001\]](#); [XU et collab. \[2010\]](#); [ZHOU et TAO \[2011\]](#) :

$$X = L + S. \quad (1.4)$$

Candes *et al.*, [CANDÈS et collab. \[2011\]](#) [WRIGHT et collab. \[2009\]](#) ont proposé une optimisation convexe pour estimer la matrice L solution du problème suivant :

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 = 0, \text{ sous la contrainte } X - L - S = 0. \quad (1.5)$$

Habituellement :

$$\lambda = \frac{1}{\sqrt{\max(I,J)}}. \quad (1.6)$$

Ceci est surprenant car on aurait pu s'attendre à ce qu'il faille choisir le bon scalaire  $\lambda$  pour équilibrer les deux termes en  $\min_{L,S} \|L\|_* + \lambda \|S\|_1 = 0$ , peut-être en fonction de leur taille relative.

$\|L\|_* = \sum_i \sigma(M)_i$  désigne la norme nucléaire de la matrice  $M$ , à savoir la somme de ses valeurs propres, et  $\|L\|_1 = \sum_{i,j} |(M)_{i,j}|$  désigne la  $l_1$ -norme de  $M$  vu comme un long vecteur dans  $\mathbb{R}^{I \times J}$ . Sous ces hypothèses minimales, cette approche est appelée Principal Pursuit Composant (*PCP*), solution qui récupère parfaitement le faible rang et les matrices parcimonieuses.

Un autre algorithme pour l'*ACP* Robuste est l'estimateur « haute répartition » pour les composantes principales : l'approche projection-poursuite revisitée **CROUX et RUIZ-GAZEN [2005]**, qui vise à trouver les directions de dispersion maximale, de manière robuste.

Dans Lin *et al.* **LIN et collab. [2010]** deux algorithmes basés sur des multiplicateurs de Lagrange augmentés (*ALM*) sont proposés. Ces deux algorithmes sont plus rapides que *ALM*, le premier algorithme est la méthode *Exact ALM (EALM)* qui a une vitesse de convergence Q-linéaire, alors que l'*APG* (Accelerated Proximal gradient) n'est en théorie que sous linéaire. Le deuxième algorithme est une amélioration de l'*EALM* appelée *méthode Inexacte ALM (IALM)*, qui converge pratiquement aussi rapidement que l'*EALM*, mais le nombre requis de SV partiels est significativement moins élevé. L'*IALM* est au moins cinq fois plus rapide que l'*APG*, et sa précision est également plus élevée.

Dans **ZHAO et collab. [2014]** un modèle *RPCA* génératif dans le cadre Bayésien est proposé en modélisant le bruit de données comme un mélange de gaussiennes (*MoG*). Le *MoG* est un approximateur universel des distributions continues, dont le modèle est capable de s'adapter à une large gamme de bruits tels que bruit Laplacien, gaussien, impulsif et toutes combinaisons de ceux-ci. L'efficacité de cette méthode a été démontrée sur des données synthétiques avec des bruits artificiels, par des modélisations de visage, et des problèmes de soustraction de fond avec des bruits réels.

Rodriguez *et Wohlberg*, dans **RODRIGUEZ et WOHLBERG [2013]** ont proposé un algorithme simple de minimisation alternée (*FPCP*), qui est approximativement deux fois plus rapide que l'algorithme *Inexact ALM* pour calculer une solution approximative au problème *PCP*. La qualité de cette solution est comparable à la solution du problème *PCP* d'origine, et elle peut être utilisée pour initialiser un algorithme plus performant, comme *Inexact ALM*.

Ma **MA [2011]** et Goldfarb *et al.* **GOLDFARB et collab. [2013]** ont proposé la version basique et accélérée de la méthode symétrique linéarisée de direction alternée (*LSADM*) pour minimiser la somme des deux fonctions convexes. La méthode basique nécessite au plus  $O(1/\epsilon)$  itérations pour une solution optimale, sachant que pour les méthodes accélérées il faut au plus  $O(1/\sqrt{\epsilon})$  itérations avec seulement une petite quantité supplémentaire d'effort de calcul à chaque itération.

### Application sous MATLAB™

Pour étudier l'intérêt de l'ACP robuste par rapport à celle classique, nous avons repris le nuage de points elliptique utilisé dans l'ACP classique, en lui appliquant la méthode *RPCA DE LA TORRE et BLACK [2001]*. Les résultats obtenus en fixant à 1 le rang de la matrice de rang faible  $L$ , sont présentés sur les figures 1.4 et 1.5.

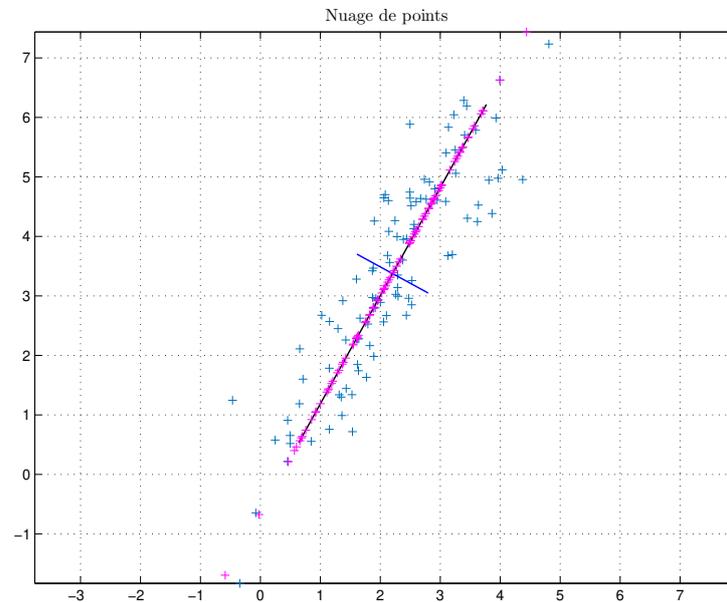


FIGURE 1.4 – ACP robuste sans point aberrant.

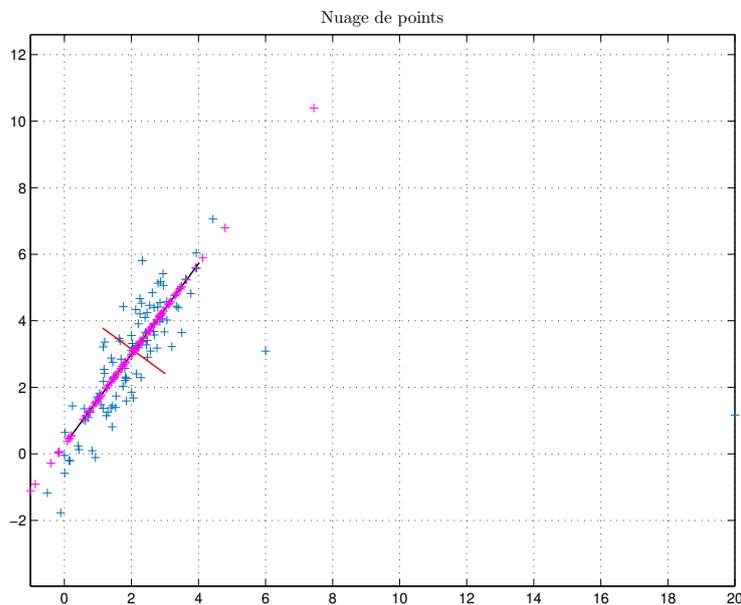


FIGURE 1.5 – ACP robuste avec points aberrants.

Sur cette figure 1.5, nous observons une petite rotation des axes principaux due aux valeurs aberrantes, mais elle est beaucoup moins importante qu'avec l'ACP classique 1.3.

Après l'illustration sur un nuage de points, nous avons appliqué l'ACP robuste sur une séquence vidéo. Dans le cas de la vidéo, le vecteur colonne de la matrice des données  $X$  représente une image (l'intensité de niveaux de gris de l'image).

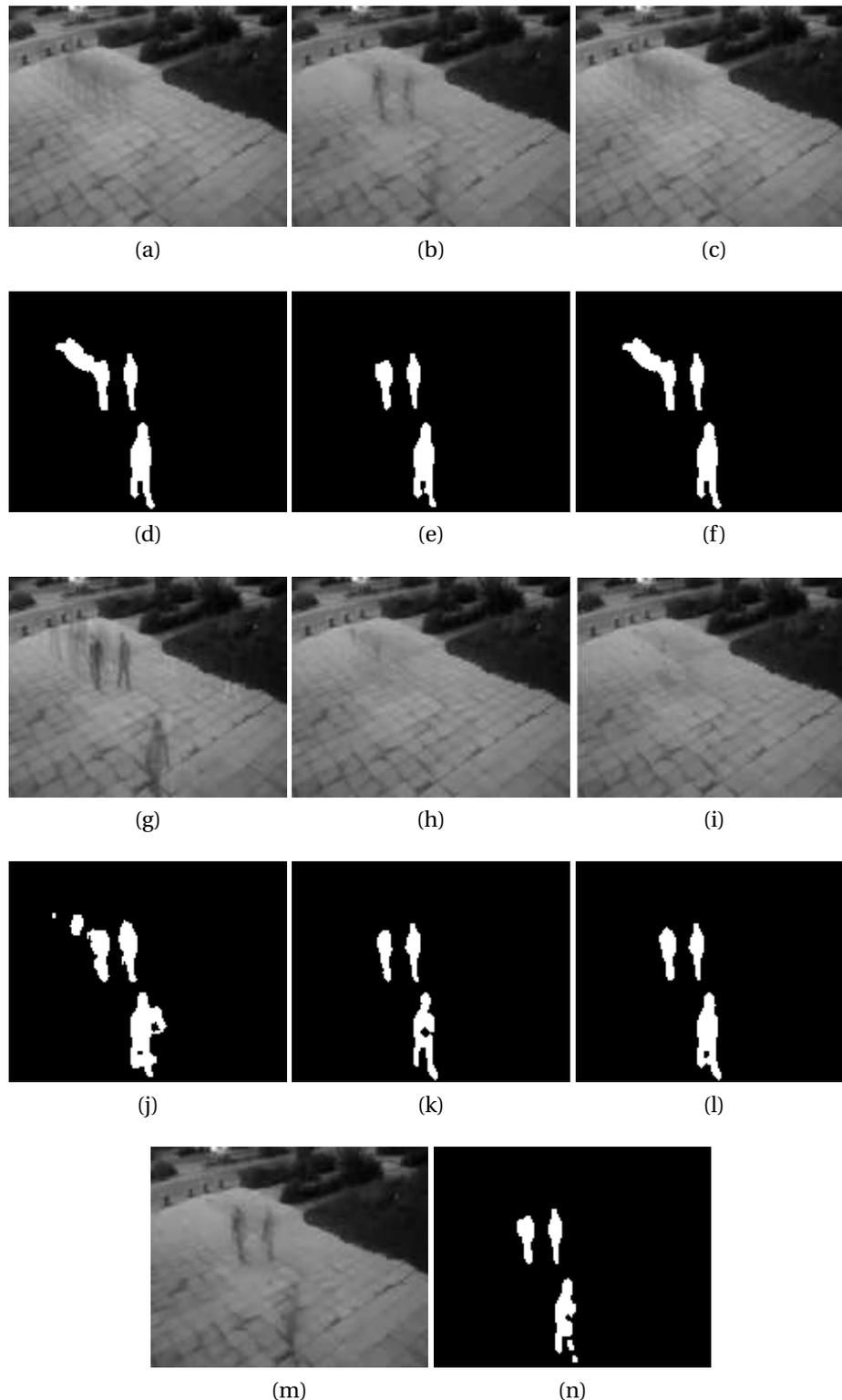


FIGURE 1.6 – ALM (a) Fond (d) Objet, EALM (b) (e), FPCP (c) (f), IALM (m) (n), RPCA (g) (j), LSADM (h) (k), MoG-ACP (i) (l).

Nous avons testé d'autres algorithmes que *RPCA*, dont notamment *LSADM*, *IALM*, *ALM*, *EALM*, *FPCP*, et *MOG-ACP* Pour cela, nous avons utilisé la librairie *LRS* [SOBRAL et collab. \[2015a\]](#).

Dans les figures suivantes, on observe à gauche l'image modèle d'arrière-plan (L, de rang 1), et à droite l'image de premier plan (S), obtenues suivant les différents algorithmes.

D'après les figures nous remarquons que les algorithmes *LSADM*, *MOG-ACP* donnent les meilleurs résultats, tandis que les autres algorithmes présentent des erreurs nettement visibles, comme des ombres dans l'arrière-plan.

## 1.6 Méthodes itératives

Les techniques itératives d'estimation du fond n'opèrent pas sur un ensemble restreint des images les plus récentes (ie. un tampon). Au lieu de cela, elles mettent à jour de manière itérative un modèle de fond, à chaque nouvelle image, à partir de cette seule image.

En conséquence, les images d'entrée du passé lointain peuvent avoir un effet sur le modèle d'arrière-plan actuel. À l'inverse des techniques non itératives, celles itératives nécessitent peu de stockage. Mais toute erreur sur le modèle de fond peut perdurer longtemps.

Dans la suite de cette section, nous décrivons certaines des techniques itératives les plus représentatives.

### 1.6.1 Différence entre images (Frame differencing)

La différence entre images utilise l'image de la vidéo à l'instant  $t - 1$  comme modèle de fond pour l'image à l'instant  $t$  :

$$|I_t - I_{t-1}| > S, \quad (1.7)$$

Avec  $I_t$  l'image à l'instant  $t$ , et  $S$  une « image » seuil (constante) à définir.

L'historique des pixels au-delà de l'image précédente n'est pas nécessaire. Il s'agit d'un avantage pour la méthode, car il n'y a ni temps de calcul prohibitif ni consommation excessive de mémoire. Cependant, comme l'historique des pixels n'est pas conservé, un arrière-plan multimodal ne peut pas être maintenu. Il est donc impossible d'éviter de détecter des objets parasites, tels que les branches d'un arbre qui se balancent. Il s'agit du problème d'ouverture décrit dans [APEWOKIN et collab. \[2007\]](#) et [TOYAMA et collab. \[1999\]](#). La comparaison réalisée dans [APEWOKIN et collab. \[2007\]](#) a confirmé la rapidité de la méthode, ainsi que ses difficultés avec les vidéos complexes.

Une approche similaire basée sur la soustraction de la première image est sujette aux mêmes problèmes. Plus précisément, chaque nouveau pixel est comparé à celui de la première image de la vidéo, et il est considéré comme arrière-plan si la différence tombe sous un certain seuil. Cette approche est très insuffisante surtout si la première image contient des objets en mouvement.

[SZOLGAY et collab. \[2011\]](#) ont proposé une nouvelle méthode pour l'extraction d'objets de premier plan dans des séquences prises par une caméra portable, avec un mouvement fort. Dans cette méthode les auteurs utilisent la différence entre images compensées pour la caméra en mouvement, améliorée avec une nouvelle estimation basée sur le noyau de la fonction de densité de probabilité des pixels de fond. Les fonctions de densité de probabilité sont utilisées pour filtrer les faux pixels de premier-plan sur la différence entre images compensées par.

L'estimation est basée sur un nombre limité de mesures; par conséquent, une sélection de points

d'échantillonnage a été introduite (méthode spatio-temporelle spécifique), et une méthode de seuillage adaptatif pour faire face à ce défi. Les objets de premier plan sont construits avec l'algorithme *DBSCAN* ((Density-Based Spatial Clustering of Applications with Noise) à partir des pixels de premier plan détectés, *DBSCAN* est un algorithme basé sur la densité dans la mesure et qui s'appuie sur la densité estimée des clusters pour effectuer le partitionnement.

## 1.6.2 Méthodes basées sur la soustraction d'une « image-fond »

Les méthodes qui suivent ont en commun de maintenir, pour chaque pixel, une estimation du niveau d'intensité du fond, et de mesurer l'appartenance d'un nouveau pixel au fond par soustraction.

Voici les étapes nécessaires :

- L'« image-fond » est initialisé avec la première image.
- Pour chaque nouvelle image, l'intensité de chaque pixel est comparée à celle du pixel correspondant de l'image-fond, par soustraction. Si la valeur absolue de cette soustraction est inférieure à un seuil donné, le pixel est classé *Fond*, et *Objet* sinon [WREN et collab. \[1997\]](#).
- Le modèle est ensuite mis à jour.

Les méthodes [MCFARLANE et SCHOFIELD \[1995\]](#); [WREN et collab. \[1997\]](#); [ZHOU et AGGARWAL \[2001\]](#) entrent typiquement dans cette catégorie.

Elles diffèrent principalement par leur manière de mettre à jour le modèle de fond (l'image-fond).

- La méthode de Filtre médian approximatif [MCFARLANE et SCHOFIELD \[1995\]](#) procède de la manière suivante : si le pixel de la nouvelle image a une intensité supérieure (respectivement inférieure) à celle de l'image-fond, alors celle-ci est incrémentée (r. décrémente).
- La méthode de détection de mouvement basique (Basic Motion Detection) [ZHOU et AGGARWAL \[2001\]](#) utilise une moyenne mobile : elle modifie l'image-fond en définissant ses niveaux comme un barycentre de son niveau courant (poids  $1 - \alpha$ ,  $\alpha$  étant petit) et de son niveau dans la nouvelle image (poids  $\alpha$ ).
- La méthode basée sur une moyenne mobile des centres de gaussiennes [CHRISTOGIANNOPOULOS et collab. \[2005\]](#); [WREN et collab. \[1997\]](#) préfère ne mettre à jour un pixel de l'image-fond, que si le pixel actuel est lui-même classé *Fond*. Ainsi, les mises à jours sont moins fréquentes et normalement plus pertinentes. Mais il ne faut pas que la première image contienne des objets, ou que le fond évolue de manière importante, car il ne pourra plus être appris.

Toutes ces méthodes sont particulièrement rapides et peu consommatrices de mémoire, et sont plutôt robustes au bruit.

Mais leur modélisation des pixels-fond étant unimodale, elles ne peuvent pas gérer correctement des fonds dynamiques.

### 1.6.3 Méthodes basées sur un mélange de gaussiennes

#### Principe et modèle

Les méthodes de la section précédente sont fondamentalement unimodales : pour chaque pixel, une seule intensité de fond est maintenue. Cela se révèle insuffisant lorsque la vidéo contient des changements dans l'arrière-plan, ou des changements brusques de luminosité. Un modèle fondé sur un ensemble de gaussiennes - donc multimodal - sera nécessairement plus souple.

Stauffer a développé l'un des plus importants algorithmes basés sur les modèles de mélange de gaussiennes *GMM* pour la soustraction de fond en temps réel [STAUFFER et GRIMSON \[1999\]](#). Il a proposé d'abord une distribution multimodale permettant une modélisation d'arrière-plan complexe, puis un algorithme pour mettre à jour ce modèle en temps réel. Chaque mode gaussien est alors supposé modéliser soit des valeurs de pixels d'arrière-plan (les valeurs les plus fréquentes), soit des valeurs de pixels de premier plan (les moins fréquentes). Ce modèle peut s'adapter aux changements d'arrière-plan, plus ou moins rapidement selon un taux d'apprentissage réglable.

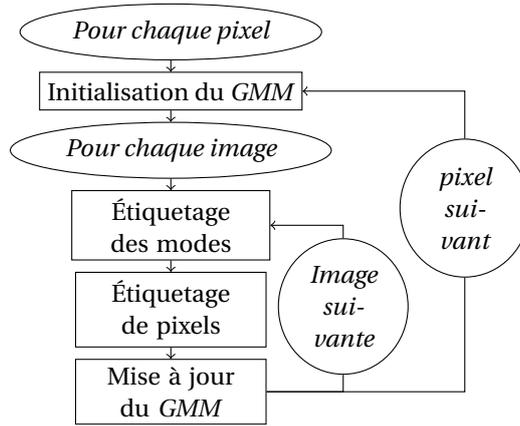


FIGURE 1.7 – Algorithme de la méthode de mélange de gaussiennes.

L'organigramme dans la figure 1.7 montre les différentes étapes de la méthode de base de soustraction du fond basée sur un mélange de gaussiennes. Le détail de chacune de ces étapes est présenté ci-dessous.

Soit  $I_t$  l'image de la vidéo  $t$  et  $p$  le pixel étudié - de coordonnées  $(i, j)$  - et  $x_t^p$  sa valeur dans l'image  $I_t$ . L'échantillon de valeurs de ce pixel particulier dans le temps est alors désigné comme suit :

$$\{x_1^p, \dots, x_T^p\} = \{I_t(i, j) : 1 \leq t \leq T\}, \quad (1.8)$$

avec  $T$  le nombre total des images de la vidéo.

Les modèles de mélange de gaussiennes *GMM* ont été largement utilisés. Ils ont la capacité d'approximation universelle, car ils peuvent modéliser toute fonction de densité de près à condition qu'ils contiennent suffisamment de composantes dans le mélange [MCLACHLAN et BASFORD \[1988\]](#). Le *GMM* associé au pixel  $p$  à l'image  $t$  est composé de  $K$  gaussiennes pondérées, et dans l'espace de couleur *RGB* chaque pixel se caractérise par son intensité dans R, G, et B. La fonction de densité d'un mélange de gaussiennes multivariées s'écrit alors :

$$f(x) = \sum_{k=1}^K w_{k,t}^p \cdot f_g(x; \mu_{k,t}^p, \Sigma_{k,t}^p), \quad (1.9)$$

avec :

- $K$  : le nombre de modes du mélange,
- $f_g(x; \mu_{k,t}^p, \Sigma_{k,t}^p)$  : la fonction de densité gaussienne du  $k^e$  mode gaussien du pixel  $p$ , à l'image  $t$ ,
- $w_{k,t}^p$  : le poids du mode  $k$ ,
- $\mu_{k,t}^p$  : son vecteur centre,
- $\Sigma_{k,t}^p$  : sa matrice de covariance.

Et avec  $f_g$  la fonction gaussienne multivariée :

$$f_g(x; \mu_{k,t}^p, \Sigma_{k,t}^p) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_{k,t}^p|}} \exp^{-1/2(x - \mu_{k,t}^p)^T \Sigma_{k,t}^p^{-1} (x - \mu_{k,t}^p)}. \quad (1.10)$$

Pour simplifier le calcul, la matrice de covariance est souvent supposée diagonale :

$$\Sigma_{k,t}^p = \sigma_{k,t}^p \mathbf{I}, \quad (1.11)$$

avec  $\mathbf{I}$  la matrice identité de taille  $3 \times 3$ .

Cela revient à supposer que les niveaux R, G, B des pixels sont indépendants et ont les mêmes variances. Ce n'est probablement pas vrai, mais cette hypothèse permet d'éviter une inversion matricielle coûteuse.

### Étape 1 : initialisation du GMM

L'idéal est d'appliquer l'algorithme *EM* (espérance-maximisation) sur une partie de la vidéo, mais on peut aussi initialiser un unique mode par pixel (de poids 1), à partir des niveaux de la 1<sup>re</sup> image.

### Étape 2 : Étiquetage des modes

Chaque mode gaussien est classé soit *Fond* soit *Objet* (premier-plan). Cette association critique est obtenue à partir d'une règle empirique : plus le mode est fréquent et précis, plus il est probable qu'il caractérise des couleurs de pixels *Fond*.

Concrètement, les  $K$  modes sont triés suivant leur niveau de priorité  $\frac{w_k}{\sigma_k}$ . Les premiers modes  $K_B$  sont alors étiquetés *Fond*. La valeur de  $K_B$  est déterminée grâce à un seuil  $T_b \in [0,1]$  :

$$K_B = \arg \min \sum_{k=1}^{K_B} w_{k,t} > T_b. \quad (1.12)$$

### Étiquetage des pixels

La troisième étape consiste à classer le pixel. Dans la plupart des méthodes, on attribue au pixel la classe du mode dont il est le plus proche, sous la contrainte :

$$|x_t^p - \mu_{k,t}^p| \leq k_p \sigma_{k,t}^p, \quad (1.13)$$

où  $k_p$  est un coefficient constant, à adapter pour chaque vidéo.

Si aucun des modes ne vérifie cette contrainte, alors le mode le moins prioritaire est remplacé par un nouveau mode gaussien centré sur l'intensité courante  $x_t^p$ , avec une variance et un poids choisis a priori [STAUFFER et GRIMSON \[1999\]](#).

### Mise à jour du GMM

Stauffer propose d'opérer de la manière suivante :

- Si la sélection d'un mode a abouti, les paramètres *GMM* sont mis à jour de manière à renforcer le mode  $i$  choisi et affaiblir les modes  $j \neq i$  rejetés :

$$w_{i,t+1} = (1 - \alpha)w_{i,t} + \alpha \quad (1.14)$$

$$\mu_{i,t+1} = (1 - \rho)\mu_{i,t} + \rho.x_{t+1}^p \quad (1.15)$$

$$\sigma_{i,t+1}^2 = (1 - \rho)\sigma_{i,t}^2 + \rho(x_{t+1}^p - \mu_{i,t+1})^2 \quad (1.16)$$

$$w_{j,t+1} = (1 - \alpha)w_{j,t}, \forall j \neq i, \quad (1.17)$$

avec  $\alpha$  un taux d'apprentissage constant et  $\rho = \alpha.f(x_{t+1}^p, \mu_i, \sigma_i)$ .

- Sinon, la dernière distribution est remplacée par un nouveau mode gaussien.

Une fois la mise à jour du modèle *GMM* effectuée, une nouvelle image peut être traitée.

### Variantes

Un avantage de l'utilisation de cette méthode est que l'arrière-plan existant n'est pas détruit lorsqu'un nouveau niveau de pixel s'intègre à l'arrière-plan.

[Zivkovic et Heijden](#). ont proposé dans [ZIVKOVIC et VAN DER HEIJDEN \[2006\]](#) un modèle *GMM* adaptatif. La méthode adapte automatiquement le nombre de gaussiennes utilisées pour modéliser un pixel donné. Cette extension réduit les exigences de mémoire des algorithmes, augmente son efficacité de calcul et peut améliorer les performances lorsque l'arrière-plan est hautement multimodal.

[Shimada et al.](#) [SHIMADA et collab. \[2006\]](#) ont proposé à leur tour une méthode qui sert à contrôler de manière dynamique le nombre de gaussiennes. Lorsque les valeurs des pixels sont constants durant un certain temps, certaines gaussiennes sont éliminées ou fusionnées. À l'inverse, le nombre de gaussiennes augmente lorsque les valeurs des pixels changent souvent.

[CARMINATI et BENOIS-PINEAU \[2005\]](#) ont proposé un algorithme de détection d'objets en mouvement, qui s'appuie sur un modèle de type mélange de gaussiennes, avec un mécanisme bayésien de maximisation de la vraisemblance. Une régularisation de Markov originale est proposée pour améliorer la détection.

#### 1.6.4 **BMOG : Modèle de mélange gaussien « boosté » avec une complexité contrôlée ( Boosted Gaussian Mixture Model with Controlled Complexity)**

MARTINS et collab. [2017] ont proposé une méthode robuste et efficace sur le plan des calculs, *BMOG*, qui augmente considérablement les performances de la méthode *GMM* STAUFFER et GRIMSON [1999]. La complexité de calcul dans *BMOG* est maintenue faible, ce qui la rend applicable en temps réel. La solution proposée dans cette méthode combine deux apports principaux : l'utilisation d'un espace couleur plus approprié que le classique *RGB*, *CIE L\*a\*b*, ainsi qu'un mécanisme de taux d'apprentissage dynamique.

#### 1.6.5 **Mélange gaussien basé-région pour la modélisation dynamique de fond *RMoG* (Region-based Mixture of Gaussians)**

VARADARAJAN et collab. [2013] ont proposé une généralisation de l'algorithme *GMM* où la relation spatiale entre les pixels est prise en compte. Fondamentalement, la classification du pixel ne dépend pas seulement de son *GMM*, mais aussi de celui de ses voisins. Les mélanges gaussiens sont associés à des régions plutôt qu'aux pixels. L'initialisation comme la mise à jour des paramètres *GMM*, sont sensiblement différentes de la méthode classique.

Un avantage de ce modèle est qu'il peut être étendu à des modèles probabilistes généraux comme les estimations de densité de noyau (*KDE*) pour la modélisation non paramétrique et aussi à d'autres familles de distributions de probabilité comme le mélange de Poisson pour des conditions de faible éclairage.

#### 1.6.6 **Méthodes utilisant le flot optique**

##### **Le flot optique**

Le concept de flot optique a d'abord été étudié dans les années 1940, puis finalement publié par le psychologue américain James J. Gibson, dans sa théorie de « l'affordance ». On appelle *flot optique* le champ de vitesses estimé à partir des variations de la luminance. Son calcul est un traitement bas niveau usuel en traitement d'images.

Les deux méthodes fondamentales de flot optique sont la méthode d'Horn et Schunck HORN et SCHUNCK [1981], et celle de Lucas et Kanade LUCAS et collab. [1981]. Elles ont beaucoup d'applications, comme l'analyse de mouvement de fluides en physique expérimentale, ou dans des traitements plus haut niveau, comme la reconstruction de scènes tridimensionnelles.

De nombreux autres algorithmes ont été proposés depuis les deux articles fondateurs. En particulier FARNEBÄCK [2003] a proposé une extension de l'algorithme de Lucas, qui permet d'estimer le mouvement de tous les pixels à partir de deux images consécutives.

## Application du flot optique à la soustraction d'arrière-plan

Cette technique de flot optique a été appliquée à notre problème, dans le but d'exploiter la continuité temporelle entre les images consécutives d'une vidéo.

Ainsi Chauhan et Krishan, CHAUHAN et KRISHAN [2013] combinent le *GMM* avec le flot optique pour l'extraction du fond.

Plus récemment CHEN et collab. [2014] ont proposé une technique similaire, plus sophistiquée, basée sur les étapes suivantes :

- Utilisation de l'arbre couvrant de poids minimal, pour définir une dissimilarité entre chaque couple de pixels de l'image, basée sur le calcul d'un chemin de continuité maximale de la couleur entre ces deux pixels. Cette distance sert à définir une notion de voisinage plus pertinente que la simple proximité géométrique.
- Estimation du flot optique utilisant l'algorithme rapide de BAO et collab. [2014], en utilisant des estimateurs robustes *M-smooth*.
- Calcul et fusion de décisions spatiale (obtenue grâce à la dissimilarité précédemment calculée) et temporelle (obtenue grâce au flot optique).

L'avantage de cette méthode est qu'elle permet une implémentation temps-réel, efficace même sur des vidéos complexes.

## 1.7 Conclusion du chapitre

Nous avons présenté les principales méthodes « classiques » de détection d'arrière plan en vidéo, ainsi que des méthodes plus récentes et performantes.

Certaines de ces méthodes sont particulièrement robustes aux différentes incertitudes présentes dans la vidéo (modification d'éclairage, arrière plan dynamique, etc.), comme les *ACP* robustes (modélisation par réduction de dimension et recherche de variance maximale). Les avantages de ces méthodes sont bien connus : modèles mathématiques fiables, exploitation aisée. Leur principal inconvénient est le temps de calcul qui peut être prohibitif, bien que certaines *ACP* robustes parviennent à éviter cet écueil.

Mais les *ACP* robustes ne sont pas seules à gérer les fonds dynamiques correctement : les mélanges de gaussiennes tirent également leur épingle du jeu. Soulignons en particulier les méthodes *BMOG* et *RMoG*, qui sont à la fois robustes et de complexité raisonnable.

Ces méthodes constituent autant de références, sur lesquelles nous nous appuyerons dans le chapitre 4 pour valider et évaluer notre méthode.

Nous présentons dans le chapitre suivant comment certaines de ces méthodes peuvent être étendues, de manière à mieux prendre en compte les incertitudes de modèles ou de données, en intégrant une modélisation par « sous-ensembles flous ». L'objectif étant de réduire encore la sensibilité au bruit ou aberrations.

## 1.8 Références

- AMATO, A., I. HUERTA, M. G. MOZEROV, F. X. ROCA et J. GONZALEZ. 2014, «Moving cast shadows detection methods for video surveillance applications», dans *Wide Area Surveillance*, Springer, p. 23–47. [9](#)
- APEWOKIN, S., B. VALENTINE, L. WILLS, S. WILLS et A. GENTILE. 2007, «Multimodal mean adaptive backgrounding for embedded real-time video surveillance», dans *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, IEEE, p. 1–6. [17](#)
- BABACAN, S. D., M. LUESSI, R. MOLINA et A. K. KATSAGGELOS. 2012, «Sparse bayesian methods for low-rank matrix estimation», *Signal Processing, IEEE Transactions on*, vol. 60, n° 8, p. 3964–3977. [13](#)
- BAO, L., Q. YANG et H. JIN. 2014, «Fast edge-preserving patchmatch for large displacement optical flow», dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 3534–3541. [23](#)
- BOUWMANS, T. 2009, «Subspace learning for background modeling : A survey», *Recent Patents on Computer Science*, vol. 2, n° 3, p. 223–234. [10](#)
- BOUWMANS, T. 2011, «Recent advanced statistical background modeling for foreground detection-a systematic survey», *Recent Patents on Computer Science*, vol. 4, n° 3, p. 147–176. [10](#)
- BOUWMANS, T. 2014, «Traditional and recent approaches in background modeling for foreground detection : An overview», *Computer Science Review*, vol. 11, p. 31–66. [10](#)
- BOUWMANS, T. et E. H. ZAHZAH. 2014, «Robust pca via principal component pursuit : A review for a comparative evaluation in video surveillance», *Computer Vision and Image Understanding*, vol. 122, p. 22–34. [8](#), [10](#)
- CALDERARA, S., R. MELLI, A. PRATI et R. CUCCHIARA. 2006, «Reliable background suppression for complex scenes», dans *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, ACM, p. 211–214. [10](#)
- CANDÈS, E. J., X. LI, Y. MA et J. WRIGHT. 2011, «Robust principal component analysis?», *Journal of the ACM (JACM)*, vol. 58, n° 3, p. 11. [13](#)
- CARMINATI, L. et J. BENOIS-PINEAU. 2005, «Gaussian mixture classification for moving object detection in video surveillance environment», dans *IEEE International Conference on Image Processing 2005*, vol. 3, ISSN 1522-4880, p. III-113–16, doi :10.1109/ICIP.2005.1530341. [21](#)
- CHAUHAN, A. K. et P. KRISHAN. 2013, «Moving object tracking using gaussian mixture model and optical flow», *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, n° 4. [23](#)
- CHEN, M., Q. YANG, Q. LI, G. WANG et M.-H. YANG. 2014, «Spatiotemporal background subtraction using minimum spanning tree and optical flow», dans *European Conference on Computer Vision*, Springer, p. 521–534. [23](#)

- CHI, E. C. et T. G. KOLDA. 2012, «On tensors, sparsity, and nonnegative factorizations», *SIAM Journal on Matrix Analysis and Applications*, vol. 33, n° 4, p. 1272–1299. [13](#)
- CHRISTOGIANNOPOULOS, G., P. B. BIRCH, R. C. YOUNG et C. R. CHATWIN. 2005, «Segmentation of moving objects from cluttered background scenes using a running average model», *SPIE Journal*, vol. 5822, p. 13–20. [18](#)
- CROUX, C. et A. RUIZ-GAZEN. 2005, «High breakdown estimators for principal components : the projection-pursuit approach revisited», *Journal of Multivariate Analysis*, vol. 95, n° 1, p. 206–226. [14](#)
- CUCCHIARA, R., C. GRANA, M. PICCARDI et A. PRATI. 2001, «Detecting objects, shadows and ghosts in video streams by exploiting color and motion information», dans *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, IEEE, p. 360–365. [10](#)
- ELGAMMAL, A., D. HARWOOD et L. DAVIS. 2000, *Non-parametric Model for Background Subtraction*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-45053-5, p. 751–767, doi :10.1007/3-540-45053-X\_48. URL [https://doi.org/10.1007/3-540-45053-X\\_48](https://doi.org/10.1007/3-540-45053-X_48). [11](#)
- ELHABIAN, S. Y., K. M. EL-SAYED et S. H. AHMED. 2008, «Moving object detection in spatial domain using background removal techniques-state-of-art», *Recent patents on computer science*, vol. 1, n° 1, p. 32–54. [10](#)
- FARNEBÄCK, G. 2003, «Two-frame motion estimation based on polynomial expansion», *Image analysis*, p. 363–370. [22](#)
- GOLDFARB, D., S. MA et K. SCHEINBERG. 2013, «Fast alternating linearization methods for minimizing the sum of two convex functions», *Mathematical Programming*, p. 1–34. [14](#)
- GUYON, C., T. BOUWMANS et E. ZAHZAH. 2012a, «Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression», dans *Pattern Recognition (ICPR), 2012 21st International Conference on*, IEEE, p. 2805–2808. [13](#)
- GUYON, C., T. BOUWMANS et E.-H. ZAHZAH. 2012b, «Moving object detection via robust low rank matrix decomposition with irls scheme», dans *Advances in Visual Computing*, Springer, p. 665–674. [13](#)
- GUYON, C., T. BOUWMANS et E.-H. ZAHZAH. 2013, «Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint», dans *Computer Vision-ACCV 2012 Workshops*, Springer, p. 315–320. [13](#)
- HORN, B. K. et B. G. SCHUNCK. 1981, «Determining optical flow», *Artificial intelligence*, vol. 17, n° 1-3, p. 185–203. [22](#)
- JEEVA, S. et M. SIVABALAKRISHNAN. 2015, «Survey on background modeling and foreground detection for real time video surveillance», *Procedia Computer Science*, vol. 50, p. 566–571. [10](#)

- LI, Y., J. YAN, Y. ZHOU et J. YANG. 2010, «Optimum subspace learning and error correction for tensors», dans *European Conference on Computer Vision*, Springer, p. 790–803. 13
- LIN, Z., M. CHEN et Y. MA. 2010, «The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices», *arXiv preprint arXiv:1009.5055*. 14
- LIU, J., J. LIU, P. WONKA et J. YE. 2012, «Sparse non-negative tensor factorization using columnwise coordinate descent», *Pattern Recognition*, vol. 45, n° 1, doi :<https://doi.org/10.1016/j.patcog.2011.05.015>, p. 649 – 656, ISSN 0031-3203. URL <http://www.sciencedirect.com/science/article/pii/S0031320311002639>. 13
- LO, B. et S. VELASTIN. 2001, «Automatic congestion detection system for underground platforms», dans *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, IEEE, p. 158–161. 10
- LUCAS, B. D., T. KANADE et collab.. 1981, «An iterative image registration technique with an application to stereo vision», . 22
- MA, S. 2011, *Algorithms for sparse and low-rank optimization : Convergence, complexity and applications*, thèse de doctorat, Columbia University. 14
- MARTINS, I., P. CARVALHO, L. CORTE-REAL et J. L. ALBA-CASTRO. 2017, *BMOG : Boosted Gaussian Mixture Model with Controlled Complexity*, Springer International Publishing, Cham, ISBN 978-3-319-58838-4, p. 50–57, doi :10.1007/978-3-319-58838-4\_6. URL [https://doi.org/10.1007/978-3-319-58838-4\\_6](https://doi.org/10.1007/978-3-319-58838-4_6). 22
- McFARLANE, N. J. et C. P. SCHOFIELD. 1995, «Segmentation and tracking of piglets in images», *Machine vision and applications*, vol. 8, n° 3, p. 187–193. 18
- McLACHLAN, G. J. et K. E. BASFORD. 1988, «Mixture models. inference and applications to clustering», *Statistics : Textbooks and Monographs, New York : Dekker, 1988*, vol. 1. 19
- PICCARDI, M. 2004, «Background subtraction techniques : a review», dans *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 4, IEEE, p. 3099–3104. 10
- POP, H. 2001, «Principal components analysis based on a fuzzy sets approach», *Mij*, vol. 1, n° 2, p. 1. 11
- RODRIGUEZ, P. et B. WOHLBERG. 2013, «Fast principal component pursuit via alternating minimization», dans *Image Processing (ICIP), 2013 20th IEEE International Conference on*, IEEE, p. 69–73. 14
- SHIMADA, A., D. ARITA et R.-I. TANIGUCHI. 2006, «Dynamic control of adaptive mixture-of-gaussians background model», dans *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*, IEEE, p. 5–5. 21
- SOBRAL, A., T. BOUWMANS et E.-H. ZAHZAH. 2015a, «Lrslibrary : Low-rank and sparse tools for background modeling and subtraction in videos», dans *Robust Low-Rank and Sparse Matrix Decomposition : Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group. 16

- SOBRAL, A., S. JAVED, S. KI JUNG, T. BOUWMANS et E.-H. ZAHZAH. 2015b, «Online stochastic tensor decomposition for background subtraction in multispectral video sequences», dans *Proceedings of the IEEE International Conference on Computer Vision Workshops*, p. 106–113. [13](#)
- STAUFFER, C. et W. E. L. GRIMSON. 1999, «Adaptive background mixture models for real-time tracking», dans *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, IEEE. [10](#), [19](#), [21](#), [22](#)
- SZOLGAY, D., J. BENOIS-PINEAU, R. MÉGRET, Y. GAËSTEL et J.-F. DARTIGUES. 2011, «Detection of moving foreground objects in videos with strong camera motion», *Pattern Analysis and Applications*, vol. 14, n° 3, p. 311–328. [17](#)
- DE LA TORRE, F. et M. J. BLACK. 2001, «Robust principal component analysis for computer vision», dans *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, IEEE, p. 362–369. [13](#), [15](#)
- TOYAMA, K., J. KRUMM, B. BRUMITT et B. MEYERS. 1999, «Wallflower : Principles and practice of background maintenance», dans *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, IEEE, p. 255–261. [17](#)
- VARADARAJAN, S., P. MILLER et H. ZHOU. 2013, «Spatial mixture of gaussians for dynamic background modelling», dans *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, IEEE, p. 63–68. [22](#)
- WREN, C. R., A. AZARBAYEJANI, T. DARRELL et A. P. PENTLAND. 1997, «Pfinder : Real-time tracking of the human body», *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, n° 7, p. 780–785. [18](#)
- WRIGHT, J., A. GANESH, S. RAO, Y. PENG et Y. MA. 2009, «Robust principal component analysis : Exact recovery of corrupted low-rank matrices via convex optimization», dans *Advances in neural information processing systems*, p. 2080–2088. [13](#)
- XU, H., C. CARAMANIS et S. SANGHAVI. 2010, «Robust pca via outlier pursuit», dans *Advances in Neural Information Processing Systems 23*, édité par J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel et A. Culotta, Curran Associates, Inc., p. 2496–2504. URL <http://papers.nips.cc/paper/4005-robust-pca-via-outlier-pursuit.pdf>. [13](#)
- ZHAO, Q., D. MENG, Z. XU, W. ZUO et L. ZHANG. 2014, «Robust principal component analysis with complex noise», dans *International Conference on Machine Learning*, p. 55–63. [14](#)
- ZHOU, G., A. CICHOCKI et S. XIE. 2012, «Fast nonnegative matrix/tensor factorization based on low-rank approximation», *IEEE Transactions on Signal Processing*, vol. 60, n° 6, doi :10.1109/TSP.2012.2190410, p. 2928–2940, ISSN 1053-587X. [13](#)

ZHOU, Q. et J. K. AGGARWAL. 2001, «Tracking and classifying moving objects from video», dans *Proceedings of IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, vol. 12, Hawaii, USA. 18

ZHOU, T. et D. TAO. 2011, «Godec : Randomized low-rank & sparse matrix decomposition in noisy case», . 13

ZIVKOVIC, Z. et F. VAN DER HEIJDEN. 2006, «Efficient adaptive density estimation per image pixel for the task of background subtraction», *Pattern recognition letters*, vol. 27, n° 7, p. 773–780. 21

# Chapitre 2

## État de l'art sur l'application des ensembles flous à la soustraction du fond en vidéo

### Sommaire

---

<b>2.1 Introduction</b> . . . . .	<b>31</b>
<b>2.2 Incertitudes et imperfections en vidéo</b> . . . . .	<b>31</b>
2.2.1 Introduction . . . . .	31
2.2.2 Sources possibles d'imperfections dans une mesure . . . . .	32
2.2.3 Mesures classiques d'incertitudes . . . . .	32
<b>2.3 Modélisation de l'incertitude par les ensembles flous</b> . . . . .	<b>33</b>
2.3.1 Ensembles flous classiques (de type-1) . . . . .	34
2.3.2 Ensembles flous basés sur un couple de fonctions d'appartenance . . . . .	36
2.3.3 Ensembles flous de type-2 . . . . .	39
2.3.4 Mesures de quantité de « flou » des ensembles flous . . . . .	41
<b>2.4 Méthodes de soustraction de fond basées sur le flou</b> . . . . .	<b>44</b>
2.4.1 Différents niveaux d'insertion du flou . . . . .	44
2.4.2 Modélisation d'arrière-plans dynamiques par un <i>GMM</i> de type <i>IVFS</i> . . . . .	46
2.4.3 La méthode <i>T2-FGMM</i> . . . . .	46
2.4.4 Illustration de la méthode <i>T2-FGMM</i> . . . . .	52
2.4.5 Critiques de la méthode <i>T2-FGMM</i> . . . . .	54
<b>2.5 Conclusion du chapitre</b> . . . . .	<b>55</b>

**2.6 Références** ..... **55**

---

## 2.1 Introduction

Dans le chapitre précédent, nous avons présenté les nombreuses imperfections pouvant affecter les vidéos (variation d'intensité lumineuses, occlusions, etc.). Ces imperfections pouvant dégrader les traitements associés à ces vidéos (extraction d'informations pertinentes, etc.), il est crucial de les prendre en compte, avec une méthode adaptée.

Parmi ces méthodes, on compte des outils statistiques classiques (*ACP* robuste, mélanges de gaussiennes, etc.), probabilistes ou descriptives.

Mais il existe de nombreuses autres méthodes et théories pour modéliser les incertitudes en mathématiques. [SUGENO \[2013\]](#) a présenté une classification des différents incertitudes modélisables à l'heure actuelle : imprécision, incomplétude, phénomènes aléatoires, non-spécificité des informations, etc. Les phénomènes aléatoires sont très bien pris en compte par la théorie des probabilités (les applications des « mélanges de gaussiennes » en traitement des images et des vidéos sont très nombreuses à ce jour). En général, l'incertitude considérée comme phénomène subjectif, peut être modélisée par des théories très différentes selon les causes de l'incertitude, le type et la quantité d'informations disponibles, les exigences de l'observateur, etc. De nombreux modèles sont associés à ces incertitudes : modèles probabilistes, théorie de l'évidence, Dempster-Shafer, ensembles flous, etc.

Les ensembles flous sont souvent associés à la modélisation de connaissances vagues (« un arbre haut », « une forte chaleur »), mais leur théorie ne se limite pas à ce type d'incertitude. Ils sont tout à fait aptes à modéliser d'autres forme d'imprécision, dans des situations bien définies, en choisissant le modèle le plus adapté (modèles flous classiques, de type-2, etc.), comme nous allons le présenter dans ce chapitre.

## 2.2 Incertitudes et imperfections en vidéo

### 2.2.1 Introduction

La modélisation d'incertitude s'est imposée dans de nombreux domaines au cours de ces dernières décennies. On peut les rencontrer dans les domaines des sciences de la décision, l'intelligence artificielle, la recherche juridique, l'économie, la science médicale, la théorie des systèmes ouverts organisationnels, la psychologie, la physique, etc.

L'incertitude en mathématique est souvent associée à une mesure, mesure classique (degré d'appartenance dans le cas des ensembles flous) ou entropie. L'imperfection d'un phénomène ou d'un processus est quand à elle associée à un mesure issue de capteur (capteur proprioceptif ou capteur extéroceptif). Lors de la présentation du résultat d'une mesure d'une grandeur physique, il est obligatoire qu'une certaine indication quantitative de la qualité du résultat soit donnée pour que ceux qui l'utilisent puissent évaluer sa fiabilité. Sans cette indication, les résultats des mesures ne peuvent être comparés, que ce soit entre eux, ou avec des valeurs de référence indiquées dans une norme ou spécification. Il est donc nécessaire qu'il y ait une procédure facilement mise en oeuvre, facile à comprendre et généralement acceptée, pour caractériser la qualité d'un résultat d'une mesure [IBWM et collab. \[1995\]](#).

### 2.2.2 Sources possibles d'imperfections dans une mesure

- Définition incomplète du mesurande.
- Imperfection de la définition du mesurande.
- Échantillonnage non représentatif : l'échantillon mesuré peut ne pas représenter le mesurande défini.
- Connaissance insuffisante des effets des conditions d'environnement sur le mesurage ou mesurage imparfait des conditions d'environnement.
- Biais dû à l'observateur pour la lecture des instruments analogiques.
- Résolution finie de l'instrument ou seuil de mobilité.
- Valeurs inexactes des étalons et matériaux de référence.
- Valeurs inexactes des constantes et autres paramètres obtenus de sources extérieures et utilisés dans l'algorithme de traitement des données.
- Approximations et hypothèses introduites dans la méthode et dans la procédure de mesure.
- Variations entre les observations répétées du mesurande dans des conditions apparemment identiques.

Ces sources ne sont pas nécessairement indépendantes, et d'autres sources peuvent contribuer à la source mesurée [ENV13005 \[1999\]](#). L'idéal serait de pouvoir associer à chaque imperfection de mesure de capteur une mesure mathématique!

Nous allons essayer maintenant d'associer imperfection de l'observation et mesure d'incertitude.

### 2.2.3 Mesures classiques d'incertitudes

Dans leurs travaux, [BLOCKLEY et GODFREY \[2000\]](#) ont caractérisé différents types d'incertitude, en les plaçant dans un espace à 3 composantes appelé « FIR », pour Flou, Incomplétude et Aléatoire. Cet espace est représenté sur la figure 2.1 (qui rend également compte des sources précédentes, avec d'autres termes tels que l'ambiguïté, le confusion, la contingence, l'indétermination et le conflits émergeant à partir des mélanges et des interactions entre ces trois attributs de base).

Cette classification en 3 composantes semble la plus pertinente à ce jour pour caractériser les types d'incertitudes altérant les mesures de phénomènes physiques.

La probabilité décrit l'incertitude de l'occurrence d'un événement, elle permet la prédiction de l'événement par la connaissance de sa fréquence relative dans un grand nombre d'expériences. Pour traiter l'incertitude stochastique, tels que les processus aléatoires, la théorie des probabilités est un outil puissant, largement répandu dans le domaine de la vision par ordinateur.

Il existe cependant d'autres formes d'incertitudes, qui se prêtent moins bien à la modélisation probabiliste. Par opposition à la probabilité, la logique floue représente l'imperfection dans le contenu informationnel de l'événement (même après la mesure, il arrive que l'on ne puisse pas établir si l'événement a eu lieu, ou non). Dans notre cas l'incertitude vient de plusieurs origines, comme les problèmes de la luminosité, les tremblements de la caméra, les occlusions entre les objets en mouvement, etc.

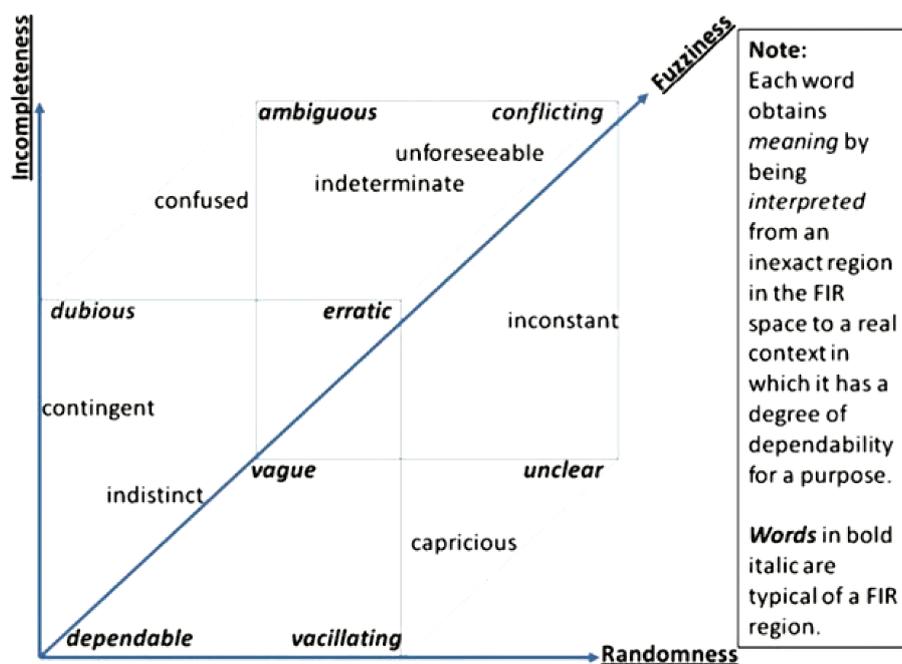


FIGURE 2.1 – Interprétations de l'incertitude dans l'espace « FIR ».

## 2.3 Modélisation de l'incertitude par les ensembles flous

La théorie des ensembles flous a été proposée par ZADEH [1965]. Cette théorie est spécifiquement conçue pour représenter mathématiquement l'incertitude et l'imprécision de données, celles-ci étant considérées dans le sens de « vague », plutôt que comme un manque de connaissances sur la valeur d'un paramètre (comme dans l'analyse de la tolérance).

La théorie des ensembles flous fournit un cadre mathématique strict, dans lequel des phénomènes conceptuels vagues peuvent être étudiés de manière précise et rigoureuse. Les ensembles flous fournissent un formalisme permettant de traiter l'imprécision intrinsèque à de nombreux problèmes. En considérant l'imprécision, la théorie des ensembles flous a fait évoluer de différentes façons de nombreuses disciplines, et constitue un pas vers un rapprochement entre la précision des mathématiques classiques et la subtile imprécision du monde réel ZADEH [1965].

Il existe de nombreuses applications des ensembles flous, par exemple, en médecine, en génie du contrôle industriel, en intelligence artificielle, en informatique, en théorie de la décision, en systèmes experts, en logique, en sciences de gestion, en recherche opérationnelle, en reconnaissance de formes, en robotique, etc.

En particulier, les ensembles flous ont été largement utilisés pour représenter l'information « Image ». L'objectif le plus important du traitement d'images floues consiste à réduire l'incertitude dans l'information d'image en « bas-niveau », comme l'imprécision due à la « pixellisation » et au bruit (thermique, de transmission des images, etc.). On peut aussi modéliser des notions vagues de types « voitures rouges » ou « petits camions » au niveau de la classification des images (traitement d'images « haut-niveau »).

Les ensembles classiques sont des ensembles avec des limites (frontières) nettes, non ambiguës. Un ensemble est une collection bien définie d'objets. Un objet dans un ensemble s'appelle un élément ou un membre de cet ensemble. Étant donné un univers de discours  $X = \{x\}$ , un ensemble net (classique)  $A$  est

défini par l'énumération de tous les éléments  $x \in X$  qui appartiennent à A :

$$A = \{x_1, \dots, x_n\}. \quad (2.1)$$

A étant un sous-ensemble de l'ensemble universel X ( $A \subseteq X$ ), on peut le caractériser par une fonction caractéristique, qui associe à tout élément  $x \in X$  un degré d'appartenance  $\mu_A(x)$  à valeur binaire :

$$\mu_A = \begin{cases} 1, & \text{si } x \in A; \\ 0, & \text{si } x \notin A. \end{cases} \quad (2.2)$$

### 2.3.1 Ensembles flous classiques (de type-1)

#### Rappels

Dans le monde réel, il existe beaucoup de connaissances « floues », soit vagues, imprécises, incertaines, ambiguës, inexactes ou encore de nature probabiliste. L'ensemble net fondé sur les deux valeurs de vérité, vraies et fausses (0 et 1), est souvent insuffisant lorsqu'on décrit le raisonnement humain. La logique floue a été introduite pour dépasser ce formalisme. Les valeurs de vérité ne sont plus restreintes à 0 et 1, elles peuvent prendre n'importe quelle valeur de l'intervalle [0,1].

Un sous-ensemble flou A est donc caractérisé par une fonction d'appartenance de ce type :

$$\mu_A : X \rightarrow [0,1], \quad (2.3)$$

où  $\mu_A(x)$  représente la mesure dans laquelle un événement  $x$  « peut être » un membre de A.

La fonction d'appartenance - ou de vérité - ainsi définie n'est pas une densité de probabilité, car la vérité probabiliste est associée à la fonction de vraisemblance (« likelihood ») d'un événement ou d'une condition, alors que la vérité floue représente une appartenance à des ensembles vaguement définis. Elle n'est donc pas soumise aux mêmes contraintes mathématiques (en particulier, son intégrale sur X ne vaut pas nécessairement 1).

La figure 2.2 montre la différence entre l'ensemble flou et l'ensemble net sachant par exemple que pour  $x = 1,5$ , le degré d'appartenance à l'ensemble net est égal à 0, alors que le degré d'appartenance à l'ensemble flou est compris entre 0 et 1 ( $\approx 0,25$ ).

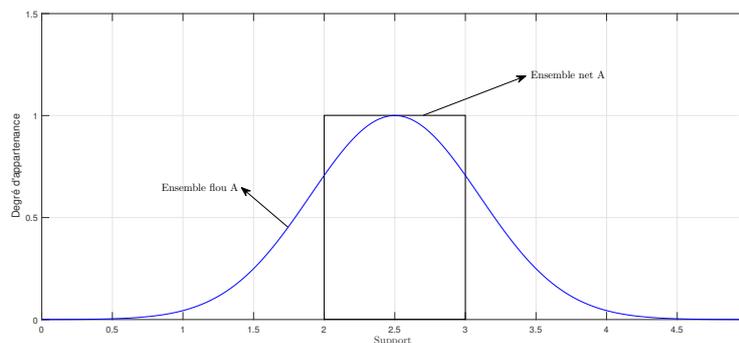


FIGURE 2.2 – Ensemble flou vs ensemble net (intervalle)

En résumé, la théorie des sous-ensembles flous est une extension de la théorie des ensembles classiques, où les éléments d'un sous-ensemble ont des degrés d'appartenance dans  $[0,1]$ .

### Alpha-coupes

Un ensemble flou est une collection d'objets avec différents degrés d'appartenance. Il est souvent utile d'examiner ses éléments qui ont au moins un degré d'appartenance égal à  $\alpha$  (c'est comme identifier les élèves d'une classe ayant obtenu une note supérieure à  $\alpha$  à un examen, ou sélectionner les montagnes d'une hauteur de plus de  $\alpha$  mètres). L'ensemble de ces éléments est un ensemble classique (net), que l'on nomme « alpha-coupe » de niveau  $\alpha$  **WIERMAN [2010]**.

À chaque  $\alpha \in ]0,1[$ , un ensemble flou  $A$  associe un ensemble net  $A^\alpha$  qui contient les éléments de l'univers  $X$  ayant un degré d'appartenance à  $A$  plus grand que  $\alpha$  :

$$A^\alpha = \{ x \in X : A(x) \geq \alpha \}. \quad (2.4)$$

Un exemple de nombre flou avec une coupe de niveau  $\alpha$  et son support est donné figure 2.3. On peut y observer la coupe  $A^\alpha = [x_1, x_2]$ , avec  $x_1$  et  $x_2$  les 2 valeurs de  $X$  correspondant aux abscisses des points d'intersection de la droite de coupe  $y = \alpha$  avec la fonction d'appartenance  $\mu_A$  **SHRESTHA et collab. [2007]**.

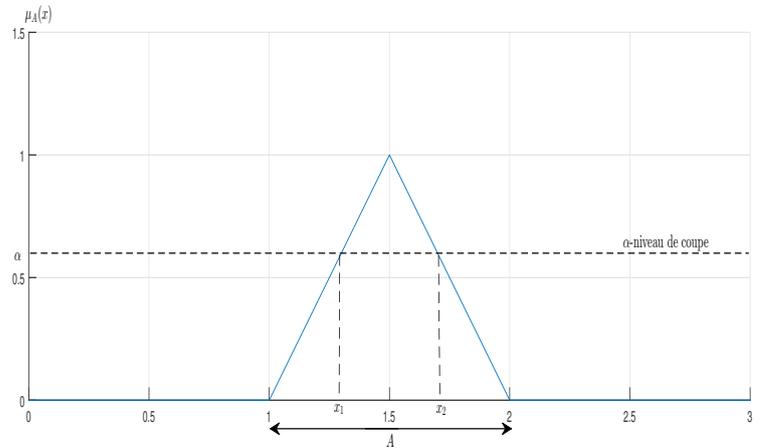


FIGURE 2.3 – Nombres flous et coupe de niveau  $\alpha$ .

La fonction d'appartenance  $\mu_A(x)$  de l' $\alpha$ -coupe  $A^\alpha$  est définie ainsi :

$$\mu_A^\alpha : X \rightarrow \{0,1\}, \quad (2.5)$$

$$\forall x \in X, \mu_A^\alpha(x) = \begin{cases} 1 & \text{si } \mu_A(x) \geq \alpha \\ 0 & \text{sinon.} \end{cases} \quad (2.6)$$

Cependant le degré d'appartenance à un ensemble flou s'exprime encore par une valeur précise (dans l'intervalle unitaire). La capacité humaine de capture d'information étant plus nuancée, il peut être judi-

cieux de développer des approches pour prendre en compte imprécision et incertitude dans les degrés d'appartenance des ensembles flous.

Une avancée importante a été l'introduction des ensembles flous intuitifs [ATANASSOV \[1986\]](#) et ensembles flous à intervalles valués (*IVFS* par la suite) pour la représentation de degrés d'appartenance imprécis. Ils ont permis une réelle avancée dans la prise de décision sous incertitudes.

Les sous-ensembles flous de type-2 constituent une alternative encore plus riche, mais aussi plus complexe à mettre en œuvre.

### 2.3.2 Ensembles flous basés sur un couple de fonctions d'appartenance

Dans cette section nous allons présenter les ensembles flous « non-standards » (*intuitifs, pythagoriciens, IVFS*) qui permettent l'expression de degrés d'appartenance imprécis, en se basant sur deux fonctions d'appartenance au lieu d'une.

Ces modèles constituent un premier pas vers les sous-ensembles flous de type-2 (ils en font d'ailleurs partie). C'est un « premier pas » en terme de simplicité de mise en œuvre, car la famille générique des sous-ensembles de type-2 généralisent les ensembles de type-1 avec des outils beaucoup plus complexes.

Leur représentation de l'incertitude est plus riche que celle des sous-ensembles flous de type-1. Ils peuvent donc être très intéressants à utiliser dans le cas de vidéos sujettes aux perturbations.

Malheureusement, cette force les rend aussi plus difficiles à utiliser et à comprendre que les autres modèles. Par conséquent, leur utilisation n'est pas encore très répandue, mais elle se développe néanmoins.

#### Ensembles flous intuitifs

Dans [ATANASSOV \[1986\]](#), l'auteur a généralisé le concept d'ensemble flou en proposant deux fonctions d'appartenance sur  $X$ , l'univers du discours.

Un ensemble flou intuitif d'Atanssov (appelé *AIFS* ou *IFS*) est défini par :

$$A = \{(x, \mu_A(x), \nu_A(x)), x \in X\}, \quad (2.7)$$

où  $\mu_A$  définit le degré d'appartenance (fonction d'appartenance) et  $\nu_A$  définit le degré de non-appartenance (fonction non-appartenance) de  $x$  à  $A$ , avec  $\mu_A : X \rightarrow [0,1]$  et  $\nu_A : X \rightarrow [0,1]$  (cf. figure 2.4), sous la condition :

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1, \forall x \in X. \quad (2.8)$$

Pour un *AIFS*  $A$  et  $x \in X$ , la fonction  $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$  mesure le degré d'indétermination de  $x$  dans  $A$ . Elle est aussi appelée indice flou intuitif, mesure d'hésitation, ou encore degré d'ignorance de  $x$  à  $A$ . Elle traduit l'incertitude de la connaissance due à un environnement conflictuel (entre l'appartenance et la non-appartenance) [ATANASSOV et GARGOV \[1989\]](#); [ATANASSOV \[1999\]](#).

On peut alors définir un ensemble flou intuitif  $A$  dans  $X$  par :

$$A = \{(x, \mu_A(x), \nu_A(x), \pi_A(x)), x \in X\}, \quad (2.9)$$

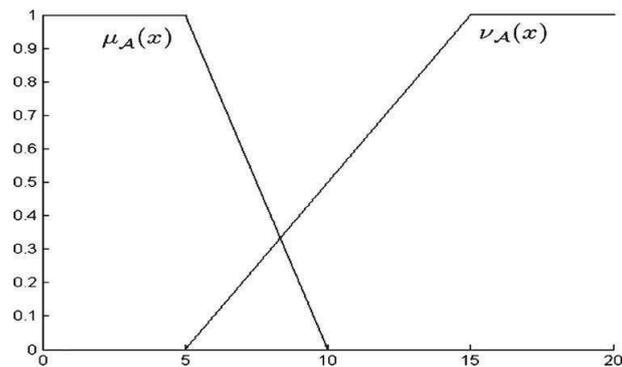


FIGURE 2.4 – Ensembles flous intuitifs avec fonction d'appartenance et de non-appartenance.

sous la contrainte :

$$\pi_A(x) + \mu_A(x) + \nu_A(x) = 1. \quad (2.10)$$

Logiquement, on pourrait estimer que l'appartenance et la non-appartenance sont complémentaires (soit  $\mu_A(x) + \nu_A(x) = 1$  ici) ; c'est en effet un postulat des sous-ensembles flous de type-1, où la non-appartenance se déduit de l'appartenance.

Mais dans le monde réel, il peut paraître difficile aux humains de fixer avec une absolue précision un degré d'appartenance  $\mu_A(x)$  (ou, de manière équivalente,  $\nu_A(x)$ ). En cas d'hésitation, il pourrait préférer s'accorder une marge d'imprécision :  $\pi(x) = 1 - (\mu_A(x) + \nu_A(x)) > 0$ . Cette part d'hésitation est donc prise, soit sur l'appartenance, soit sur la non-appartenance, sans distinction ; c'est justement en cela qu'elle traduit une ambiguïté.

Soit un élément  $x \in X$ , la condition  $\mu_A(x) + \nu_A(x) \leq 1$  sur les *AIFS* suggère que si l'une des deux fonctions d'appartenance / non-appartenance a une grande valeur (proche de 1), alors l'autre fonction devrait avoir une très petite valeur (proche de 0) [DESPI et collab. \[2013\]](#). Mais il est possible que les deux fonctions aient des petites valeurs, c'est-à-dire que le degré d'adhésion et le degré de non-adhésion soient assez négligeables, au profit de la part d'hésitation.

C'est ce que l'on peut observer sur l'illustration de l'appartenance d'un élément  $x$  à un *AIFS*  $A$  (cf. figure 2.5). La diagonale représente les degrés d'appartenance modélisés par les ensembles flous ordinaires, alors que la surface grise représente les appartenances possibles des *AIFS*. D'après Atanassov les fonctions d'appartenance et de non-appartenance doivent avoir leur somme inférieure ou égale à 1 pour chaque élément de l'univers du discours.

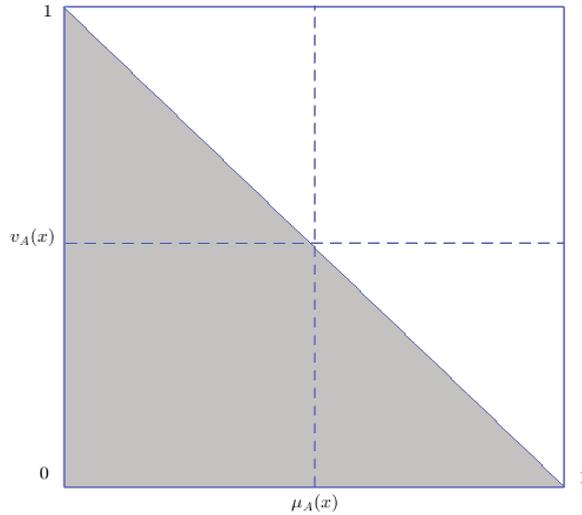


FIGURE 2.5 – Ensemble flou intuitif d'Atanassov  $\mu_A(x) + \nu_A(x) \leq 1 \forall x \in X$ .

### Ensembles flous intervalles (IVFS)

À l'instar des ensembles flous intuitifs, les ensembles flous intervalles (en anglais, *Interval-Valued Fuzzy Sets*, IVFS) ont été proposés pour pallier la trop grande précision des degrés d'appartenance des sous-ensembles flous de type-1 [HISDAL \[1981\]](#).

Leur principe consiste à encadrer les fonctions d'appartenance des ensembles flous de type-1, par deux fonctions d'appartenance : une qui borne inférieurement les degrés d'appartenance possibles, l'autre supérieurement.

Soit un IVFS  $\tilde{A}$ , il est donc défini par le couple de fonctions d'appartenance  $\{\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x)\}$ .

Nous n'allons pas plus loin dans la description de ce modèle, car il a été montré qu'il était équivalent à celui des ensembles flous intuitifs [ATANASSOV \[1999\]](#); [CHAIRA et RAY \[2008\]](#); [DESCHRIJVER et KERRE \[2007\]](#).

En particulier, les fonctions d'appartenance des deux modèles vérifient la relation suivante :

$$\begin{aligned} \underline{\mu}_{\tilde{A}}(x) &= \mu_{\tilde{A}}(x); \\ \overline{\mu}_{\tilde{A}}(x) &= 1 - \nu_{\tilde{A}}(x). \end{aligned} \tag{2.11}$$

### Ensembles flous Pythagoriciens

L'ensemble flou « pythagoricien » (PFS) a été introduit par Yager [YAGER \[2013, 2014\]](#). C'est un modèle proche de celui des ensembles flous intuitifs, avec un peu plus de degrés de liberté.

Il est caractérisé par un degré d'appartenance et un degré de non-appartenance dont la somme des carrés est inférieure ou égale à 1. C'est une généralisation du nombre flou intuitif (IFS), puisque ceux-ci satisfont nécessairement cette contrainte « pythagoricienne ».

Yager [YAGER \[2013, 2014\]](#), a donné trois représentations de base pour la catégorie des nombres pythagoriciens :

- Le nombre flou  $(a, b)$  satisfaisant les conditions  $a \in [0, 1]$ ,  $b \in [0, 1]$  et  $a^2 + b^2 \leq 1$ .
- Les coordonnées polaires  $(r, \theta)$  satisfaisant les conditions  $r \in [0, 1]$  et  $\theta \in [0, \pi/2]$ .

— Il faut que  $(r,d),\theta \in [0,\pi/2]$ , et la direction de la paire d'engagement  $d = \frac{\pi-2\theta}{\pi}$ .

Avec la relation :  $a^2 + b^2 = r^2$ ,  $a = r \cos(\theta)$ ,  $b = r \sin(\theta)$ .

Un  $PFS\beta$  est défini par :

$$\beta = \{ \langle x, \beta(\mu_\beta(x), \nu_\beta(x)) \rangle, x \in X \}, \quad (2.12)$$

où  $\mu_\beta : X \rightarrow [0,1]$  représente le degré d'appartenance, et  $\nu_\beta : X \rightarrow [0,1]$  représente le degré de non-appartenance de  $x \in X$  à  $\beta$ , sous la contrainte :

$$\left( (\mu_\beta(x))^2 + (\nu_\beta(x))^2 \right) \leq 1. \quad (2.13)$$

À la différence des nombres flous intuitifs, le degré d'indétermination (hésitation) de  $x$  à  $\beta$  est défini par :

$$\pi_\beta(x) = \sqrt{1 - (\mu_\beta(x))^2 - (\nu_\beta(x))^2}. \quad (2.14)$$

Sur la figure 2.6, on peut observer la différence entre la contrainte de complémentarité des nombres pythagoriciens et celle des nombres flous intuitifs. En particulier l'espace modélisant l'incertitude est plus vaste avec les  $PFN$  qu'avec les  $IFN$ . Cela souligne l'intérêt des  $PFS$  en terme de gestion de l'incertitude par rapport aux  $IFN$ .

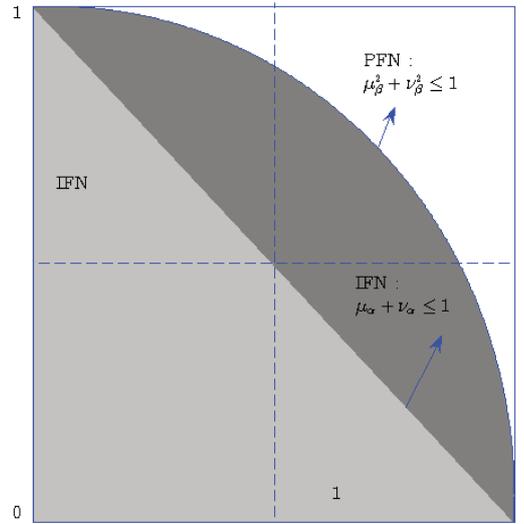


FIGURE 2.6 – Les espaces d'incertitude des  $PFN$  et  $IFN$ .

### 2.3.3 Ensembles flous de type-2

Les ensembles flous de type-2 ont été proposés dans le but de modéliser et de minimiser les effets des incertitudes dans les systèmes de logique floue.

Leur représentation de l'incertitude est encore plus développée que celle des modèles précédents. Ils ont été par exemple utilisés dans un premier dans des contrôleurs flous de type -2 R.BOUKEZZOULA et L.FOULLOY [2016], et la mise au point de classifieurs performants basés sur les « fuzzy c-means » HWANG et RHEE [2007]. Mais leur caractérisation et leur manipulation sont plus complexes que les ensembles flous classiques, ce qui pénalise beaucoup leur utilisation, notamment en vidéo.

Nous les présentons néanmoins car leur formalisme élégant généralise tous les modèles précédents.

### Définition

L'ensemble flou ordinaire utilisé dans le traitement d'images est souvent un nombre flou. Un ensemble flou définit la représentation du sens de l'imprécision associée A (A variable linguistique dans une langue naturelle [DELEPOULLE et collab. \[2011\]](#)). Toutefois, il est impossible de dire quelle fonction d'appartenance est la plus adaptée au problème considéré. Soit X l'univers de discours, dans l'ensemble flou de type-1, la fonction d'appartenance est un nombre, alors que dans T2FS, la fonction d'appartenance est elle-même un ensemble flou [HOSSEINI et collab. \[2010\]](#). Le T2FS a été introduit par Zadeh en 1975 et a été étendu (entre autres) par Karnik et Mendel en 1998 [KARNIK et collab. \[1999\]](#).

Karnik *et al.* définissent un ensemble flou  $\tilde{A}$ , à l'aide d'une fonction caractéristique  $\mu_{\tilde{A}} : X \times [0,1] \rightarrow [0,1]$  à deux paramètres :

$$\tilde{A} = \{(x, u), \mu_{\tilde{A}}(x, u), \forall x \in X, \forall u \in J_x \subseteq [0,1]\}. \quad (2.15)$$

Elle permet d'associer à chaque valeur  $x \in X$ , non pas un singleton - trop précis - dans  $[0,1]$  (comme les ensembles flous de type-1), mais un nombre flou de domaine  $[0,1]$ . Donc, pour le  $x$  considéré,  $\mu_{\tilde{A}}$  associe à tout élément  $u$  de ce domaine, un degré d'appartenance - cette fois-ci singleton - de  $[0,1]$ .

Karnik *et al.* définissent en outre deux fonctions d'appartenance plus simples pour décomposer les ensembles flous de type-2 :

- la *fonction d'appartenance primaire* : à chaque  $x \in X$ , elle associe l'intervalle minimal  $J_x \subseteq [0,1]$  qui contient l'ensemble des degrés d'appartenance  $u$  tels que  $\mu_{\tilde{A}}(x, u) > 0$  : le support du nombre flou de type-1 associé à  $x$  par  $\tilde{A}$ ;
- la *fonction d'appartenance secondaire* : à chaque  $x \in X$ , elle associe la fonction d'appartenance de l'ensemble flou de type-1 associé à  $x$  par  $\tilde{A}$ .

On peut parler de manière imagée de « fonctions verticales », car elles expriment les caractéristiques de  $\tilde{A}$  pour une tranche verticale, c'est-à-dire pour un  $x \in X$  donné.

Sur la figure 2.7, on observe en noir une fonction d'appartenance secondaire.

Enfin, Karnik *et al.* désignent la région recouvrant l'ensemble des intervalles  $J_x$  définis sur X (les fonctions primaires de  $\tilde{A}$ ) par le terme d'empreinte d'incertitude, ou FOU en anglais (pour *footprint of uncertainty*) :

$$\text{FOU}(\tilde{A}) = \bigcup_{x \in X} J_x. \quad (2.16)$$

On peut l'observer en gris sur la figure 2.7.

La surface de cette empreinte mesure la quantité de « flou » des fonctions d'appartenance primaires.

Les fonctions qui délimitent cette zone, sont des fonctions d'appartenance floues de type-1. On peut les considérer comme des fonctions « horizontales », qui rendent compte de l'imprécision des fonctions

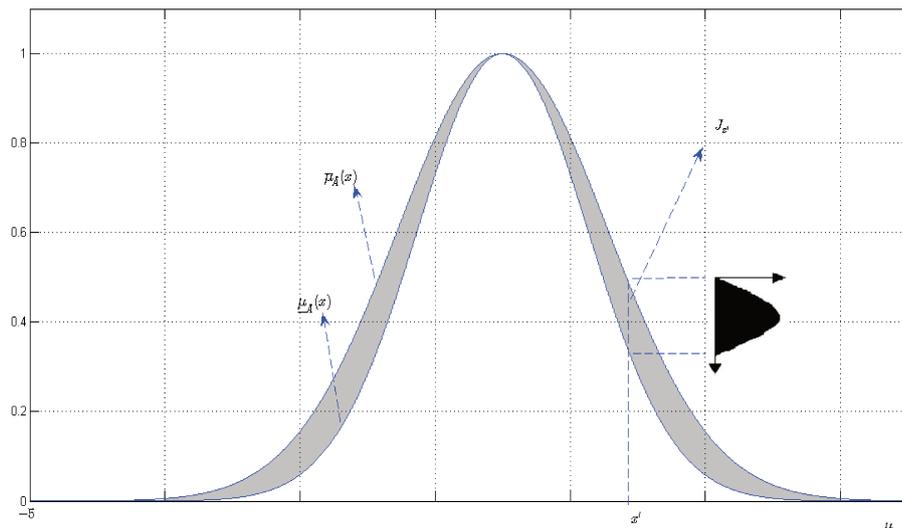


FIGURE 2.7 – Fonction d'appartenance secondaire d'un ensemble flou de type-2 pour  $x = x'$ .

d'appartenance primaire, de manière globale :

$$\begin{aligned} \overline{\mu}_{\tilde{A}}(x) &= \overline{\text{FOU}(\tilde{A})}(x) & \forall x \in X; \\ \underline{\mu}_{\tilde{A}}(x) &= \underline{\text{FOU}(\tilde{A})}(x) & \forall x \in X. \end{aligned} \tag{2.17}$$

### Liens aux autres modèles

Les ensembles flous de type-2 ont été proposés dans le but de généraliser les ensembles flous de type-1, en remplaçant les degrés d'appartenance singletons par des ensembles flous de type-1.

Les nombres intuitifs et *IVFS*, qui étendent eux-mêmes les ensembles flous de type-1, sont également des cas particuliers d'ensembles flous de type-2. On obtient en effet l'équivalence, en contraignant les fonctions d'appartenance secondaires à associer à tout  $x \in X$  un intervalle net de  $[0,1]$  (au lieu d'un ensemble flou de type-1). Toute l'incertitude est alors caractérisée par la *FOU*.

Les nombres pythagoriciens s'inscrivent également dans ce cadre.

Pour conclure, en incorporant l'incertitude sur la fonction d'appartenance dans la théorie des ensembles flous, les ensembles flous de type-2 permettent de répondre à une des critiques adressée aux ensembles flous de type-1 : de reporter l'arbitraire des bornes d'un ensemble net (un intervalle par exemple) sur les degrés d'appartenance.

On note qu'en l'absence d'incertitude de type-2 (*FOU* nulle), alors un ensemble flou de type-2 se réduit à un ensemble flou de type-1.

### 2.3.4 Mesures de quantité de « flou » des ensembles flous

Comme nous avons vu précédemment, l'incertitude en vidéo provient de plusieurs imperfections, telles que la luminosité, la caméra, l'occlusion entre deux objets en mouvement, etc. Une manière de la prendre en compte est d'utiliser des ensembles flous. Il devient alors intéressant d'évaluer leur quantité de

fou, notamment sous la forme d'une sorte de distance par rapport à un ensemble « moins fou » (soit un ensemble net, pour un sous-ensemble fou de type-1; ou un ensemble fou de type-1, pour un ensemble fou de type-2). C'est ce que permet l'entropie.

### Entropies d'un ensemble fou de type-1

Le terme d'entropie d'un ensemble fou simple A sur un ensemble universel fini  $X = \{x_1, \dots, x_i, \dots, x_q\}$  a d'abord été introduit par De Luca et Termini **DE LUCA et TERMINI [1972]** :

$$H(A) = \sum_{i=1}^q \frac{S_q(\mu_A(x_i))}{q \cdot \ln 2}. \quad (2.18)$$

Rappelons que la fonction d'entropie de Shannon est :

$$S_q(\mu_A(x_i)) = -\mu_A(x_i) \cdot \ln(\mu_A(x_i)) - (1 - \mu_A(x_i)) \cdot \ln(1 - \mu_A(x_i)). \quad (2.19)$$

De Luca et Termini ont également formulé les exigences axiomatiques qu'une mesure d'entropie doit respecter.

Yager **YAGER [1979]** et Kaufmann **KAUFMANN [1975]** ont proposé d'autres mesures possibles de l'entropie, inspirées de l'entropie de Shannon et en utilisant la mesure de la distance. Le résultat principal est le degré de fou d'un ensemble fou A qui reflète le degré d'ambiguïté en mesurant la distance entre A et son plus proche ensemble ordinaire  $A^{0.5}$ . Il est défini comme suit (voir **YAGER [1979]**) :

$$H(A) = \frac{2 \cdot d(A, A^{0.5})}{q^{1/q}}, \quad (2.20)$$

où  $d(A, A^{0.5})$  désigne la distance (distance de type  $L_p$ ) entre A et son plus proche ensemble ordinaire  $A^{0.5}$  :

$$d(A, A^{0.5}) = \sqrt[p]{\sum_{i=1}^n |A - A^{0.5}|^p}. \quad (2.21)$$

La valeur  $p$  positive permet en particulier de contraindre  $H(A)$  dans  $[0, 1]$ . Son choix détermine la distance utilisée. Par exemple,  $p = 1$  correspond à une distance de Hamming généralisée (indice linéaire de fou  $\gamma(A)$ ), tandis que  $p = 2$  correspond à une distance euclidienne (indice quadratique de fou).

Si les images numériques sont considérées comme des sous-ensembles flous, nous pouvons alors utiliser l'extension de la définition de l'indice linéaire de fou proposé par Pal **PAL et BEZDEK [1994]**. Pour une image  $M \times N$  sous-ensemble  $A \subseteq X$  avec  $G$  niveaux de gris ( $g \in [0, G - 1]$ ), l'histogramme  $h(g)$  et la fonction d'appartenance  $\mu_A(g)$ , l'indice de fou de l'image peut alors être défini comme suit **PAL et BEZDEK [1994]** :

$$\gamma(A) = \frac{2}{M \cdot N} \sum_{g=0}^{G-1} [h(g) \cdot \min(\mu_A(g), 1 - \mu_A(g))]. \quad (2.22)$$

### Entropies sur les IVFS

Burillo **BURILLO et BUSTINCE [1996]** a présenté une étude intéressante par rapport à l'entropie d'un IVFS A dans X. La définition de l'entropie est basée sur les contraintes suivantes :

- L'entropie est nulle lorsque l'ensemble est un *FS*.
- L'entropie est maximale si l'ensemble est totalement indéterminé.
- L'entropie d'un *IVFS* est égale à son complément respectif.
- Si le degré d'appartenance et le degré de non-appartenance de chaque élément augmentent, cet ensemble devient flou et donc l'entropie devrait diminuer.

La première condition de définition montre clairement que l'entropie mesure une quantité d'information qui est distincte du degré de flou. Elle le complète en incorporant le degré d'imprécision de l'ensemble flou considéré.

Parmi toutes les entropies floues possibles caractérisés par Burillo [DELEPOULLE et collab. \[2011\]](#), la mesure suivante est la plus communément retenue :

$$Ind(A) = \sum_{i=1}^q \delta_A(x_i). \quad (2.23)$$

Cette mesure est la même que le degré d'indétermination des *IVFS* (noté  $\Phi$ -ensemble flous) proposé par [SAMBUC \[1975\]](#) :

$$Ind(A) = \sum_{i=1}^q \delta_A(x_i) = \sum_{i=1}^q \left[ \bar{\mu}(x_i) - \underline{\mu}(x_i) \right]. \quad (2.24)$$

Ainsi, l'entropie d'*IVFS* identifie le degré d'indétermination (Tizhoosh le nomme « ultrafuzziness » d'un *IVFS*). Un ensemble flou *IVFS* peut donc représenter les incertitudes dans la fonction d'appartenance d'un *FS*. Donc, si dans un *FS*, les données d'entrée contiennent du bruit, le transfert de cette incertitude sur la fonction incertitude de degré d'appartenance conduit à la construction d'un *FS* imprécis (ou *IVFS*) [WANG et ACAR \[1999\]](#).

Pour une image de dimensions  $M \times N$  à  $G$  niveaux de gris, Tizhoosh [TIZHOOSH \[2005\]](#) propose l'indice (linéaire) « ultrafuzziness » d'un *IVFS*  $A$  suivant :

$$\Gamma(A) = \frac{1}{M.N} \sum_{x=0}^{G-1} [h(x) \cdot (\bar{\mu}_A(x) - \underline{\mu}_A(x))], \quad (2.25)$$

avec  $h$  la fonction histogramme de l'image.

Bigand *et al.* [BIGAND et COLOT \[2010\]](#) ont proposé une version légèrement différente, qui permet d'estimer l'incertitude sur l'appartenance des pixels de l'image à la gamme de gris  $g$  :

$$\begin{aligned} \Gamma(g) &= \frac{1}{M.N} \sum_{x=0}^{G-1} \left[ h(g) \cdot (\bar{\mu}_x(g) - \underline{\mu}_x(g)) \right] \\ &= \frac{1}{M.N} \sum_{x=0}^{G-1} \left[ h(g) \cdot (|\mu(g; x, \sigma)|^{1/\alpha} - |\mu(g; x, \sigma)|^\alpha) \right], \end{aligned} \quad (2.26)$$

où  $\bar{\mu}_x$  et  $\underline{\mu}_x$  dénotent le couple de fonctions d'appartenance d'un *IVFS* gaussien centré sur  $x$ .

## 2.4 Méthodes de soustraction de fond basées sur le flou

Malgré leur intérêt pour faire face aux bruits et autres sources d'incertitude, l'utilisation de modèles flous n'est pas extrêmement répandue dans le domaine du traitement vidéo.

Nous nous sommes rapidement focalisés sur la méthode [BOUWMANS et EL BAF \[2010\]](#); [EL BAF et collab. \[2008b, 2009\]](#), mais nous allons d'abord présenter un bref état de l'art des alternatives floues à cette méthode de soustraction de fond.

### 2.4.1 Différents niveaux d'insertion du flou

Les concepts flous ont été introduits à différents niveaux du processus général de soustraction d'arrière-plan, comme nous le rappelons ci-dessous [BOUWMANS \[2012\]](#).

#### Modélisation d'arrière-plan flou

De nombreuses approches récentes se fondent sur un modèle d'arrière-plan multimodal. Le modèle habituellement utilisé est le mélange gaussien [WREN et collab. \[1997\]](#). Les paramètres sont souvent initialisés à l'aide d'une séquence d'apprentissage, qui contient des données insuffisantes ou bruitées. Donc, les paramètres ne sont pas bien déterminés.

Dans ce contexte, les modèles de mélange gaussien flou de type-1 ou type-2 [BOUWMANS et EL BAF \[2010\]](#); [EL BAF et collab. \[2008b, 2009\]](#) permettent de prendre en compte cette imprécision.

[SUBUDHI et collab. \[2016\]](#) ont exploité un modèle flou de ce type, allié à une technique de segmentation basée sur l'analyse des textures et de la chrominance, pour proposer une méthode particulièrement apte à détecter les objets mobiles, tout en les séparant de leurs ombres.

#### Détection de premier plan flou

Cette catégorie de méthodes n'introduisent pas explicitement d'imprécision dans leurs données ou modèles - c'est-à-dire sous forme d'ensembles flous -, mais elles mesurent l'appartenance des pixels au premier plan, grâce à des degrés flous, obtenus en normalisant des mesures de similarité (ou dissimilarité) dans  $[0,1]$ .

Le modèle d'arrière-plan peut être unimodal, comme dans [SIGARI \[2008\]](#); [SIGARI et collab. \[2008\]](#), où l'intensité de chaque pixel-fond est estimée à l'aide d'une moyenne mobile. L'appartenance de chaque pixel à l'ensemble flou *Fond*, est ensuite évaluée à l'aide d'une dissimilarité proportionnelle à la distance entre son niveau et celui du pixel-fond. Un post-traitement par filtre passe-pas permet ensuite de lisser le résultat.

Le modèle peut aussi être multimodal, comme la modélisation de fond avec mesure de confiance proposée par [ROSELL-ORTEGA et collab. \[2010\]](#). [ZHANG et XU \[2006\]](#) et [EL BAF et collab. \[2008a\]](#) ont étendu cette approche en caractérisant des dissimilarités de différentes natures, entre les pixels et le pixel-fond : ces multiples dissimilarités portent à la fois sur la couleur des pixels, mais aussi sur la texture (grâce à des motifs binaires locaux). Elles sont ensuite agrégées par une intégrale de Choquet, pour obtenir au final un degré d'appartenance à l'ensemble flou *Fond*. [LU et collab. \[2014\]](#) utilise le même principe mais en se servant des motifs binaires locaux uniformes (*ULBP*) pour plus de robustesse, sachant que l'opérateur *ULBP*

améliore l'invariance de rotation du code de modèle binaire local (*LBP*) et quantifie les statistiques d'occurrence de modèles invariants de rotation individuels correspondant à certaines micro-caractéristiques de l'image. La détection de premier plan flou est plus robuste aux changements d'éclairage et aux ombres que la détection de premier plan classique.

**ZI-LONG et collab. [2013]** ont proposé une méthode floue adaptative pour les vidéos de scène complexe. Dans cette méthode, un contrôleur flou Takagi-Sugeno-Kang (*TSK*) est utilisé comme estimateur, et les paramètres de la partie prémisse et de la partie conséquence du système flou sont optimisés en combinant l'optimisation de l'essaim de particules (*PSO, particle swarm optimization* en anglais) avec la récursivité de l'estimateur des moindres carrés (*RLSE*). Afin d'estimer efficacement l'arrière-plan, les échantillons de premier plan sont interprétés comme des valeurs aberrantes par rapport aux échantillons d'arrière-plan, et une méthode de séparateur de valeurs aberrantes est mise au point. Une fois les valeurs aberrantes supprimées, les résultats obtenus sont utilisés pour former l'estimateur flou.

**CHIRANJEEVI et SENGUPTA [2014]** ont proposé un algorithme pour la détection d'objets en mouvement dans des conditions de fond dynamiques difficiles. Ils utilisent un ensemble de mesures de similarité multi-aléatoires agrégées floues appliquées sur plusieurs modèles correspondant à des arrière-plans multimodaux. L'algorithme est enrichi d'une stratégie d'initialisation de modèle prise en charge par le voisinage pour une convergence plus rapide, avec une maintenance de modèle d'arrière-plan pilotée par une mesure d'agrégation floue au niveau du modèle et qui assure une plus grande robustesse. La mise à jour du modèle et la décision de classification de premier plan / arrière-plan sont basées sur l'ensemble des intégrales floues.

### Maintenance d'arrière-plan flou

L'idée est de mettre à jour le modèle d'arrière-plan en fonction des degrés d'appartenance flous du pixel aux classes d'arrière-plan et de premier plan, ces degrés flous provenant d'une détection floue. Cette maintenance d'arrière-plan adaptative floue permet de gérer de manière robuste les changements d'éclairage et les ombres.

**KIM et KIM [2012]** puis **GUTTI et SHANKAR [2014]** ont présenté des algorithmes de soustraction d'arrière-plan exploitant ce principe de mise à jour floue. Ils se servent tous deux d'un histogramme de couleur floue (*FCH* en anglais), dans le but d'atténuer les variations de couleur générées par les mouvements du fond tout en mettant en évidence les objets en mouvement.

Ces méthodes gèrent de manière efficace le bruit de fond des scènes comportant des textures dynamiques (temporelles).

### Post-traitement flou

Un moteur d'inférence floue peut être appliqué sur les masques de premier plan courant et antérieurs, pour bien détacher l'objet en mouvement de l'arrière-plan dynamique. **MANJULA et SIVABALAKRISHNAN [2010]** ont proposé une méthode de ce type, qui utilise utilisant la caractéristique de couleur et un filtrage médian **BOUWMANS [2012]**.

## 2.4.2 Modélisation d'arrière-plans dynamiques par un *GMM* de type *IVFS*

Parmi toutes ces méthodes, nous nous sommes vite tournés vers la catégorie intégrant le flou au niveau du modèle de fond, et plus précisément vers celles utilisant un modèle fond multimodal, de type *GMM* imprécis. Ces méthodes ont en effet l'avantage d'être à la fois simples, rapides, robustes et performantes.

### Méthodes fondées sur un *GMM* imprécis

Le modèle de mélange gaussien *GMM* est la technique la plus usitée dans la modélisation de l'arrière-plan en vidéo. Le *GMM* est souvent d'abord initialisé par l'algorithme espérance-maximisation (*EM*), à partir d'une séquence d'images. Puis il est mis à jour itérativement, à chaque nouvelle image, souvent sur un principe de moyenne mobile.

Toutefois, en raison de données insuffisantes ou corrompues dans la séquence initiale d'apprentissage ou de mise à jour, le *GMM* ne reflète pas exactement la répartition sous-jacente des observations. Il peut sembler problématique d'utiliser des probabilités - qui sont elles-mêmes des nombres réels précis - pour modéliser des *GMM* à partir de données ou de paramètres incertains.

Dans un autre cadre applicatif, Zeng et al. [ZENG et collab. \[2008\]](#) et [ZENG et LIU \[2006b, 2007\]](#), ont proposé des classifieurs basés sur un *GMM* « flou ». Plus précisément, ils ont introduit un modèle d'incertitude dans le *GMM*, en remplaçant les fonctions d'appartenance gaussiennes par des *IVFS*. Ils ont par ailleurs proposé un mécanisme de mise à jour de ce modèle imprécis (ce sont des méthodes itératives). La première méthode construit ensuite la décision à l'aide d'un *GLM* (Modèle Linéaire Généralisé) et d'un *HMM* (Modèle de Markov Caché). La seconde méthode exploite plusieurs classifieurs, dont un classifieur bayésien.

Plus récemment, El Baf *et al.* [BOUWMANS et EL BAF \[2010\]](#); [EL BAF et collab. \[2008b, 2009\]](#) ont proposé un modèle semblable, en l'appliquant cette fois-ci au problème de soustraction de fond, et en proposant une procédure de décision simplifiée, inspirée de celle proposée par Stauffer dans le cas non flou [STAUFFER et GRIMSON \[1999\]](#). C'est la méthode *T2-FGMM*, qui devient alors capable d'éviter les problèmes classiquement rencontrés (fond dynamique, variation de luminosité, occlusion, etc.).

C'est cette méthode que nous avons choisi d'étudier, en particulier car elle est à la fois relativement récente et particulièrement rapide.

## 2.4.3 La méthode *T2-FGMM*

Après une description du modèle *GMM* flou exploité dans cette méthode, nous enchaînons avec le détail de son algorithme.

### Modèle *GMM* de type *IVFS*

L'algorithme proposé se base sur un *GMM* classique, dans lequel le flou est inséré au moment opportun, c'est-à-dire à l'étape **Étiquetage des pixels** (cf. sous-section [2.4.3](#)).

Le *GMM* associé au pixel  $p$  de l'image  $t$  est composé de  $K$  gaussiennes multivariées pondérées (cf. équations 1.9 et 1.10). Il permet de modéliser la distribution de sa couleur dans l'espace *RGB* (chaque pixel étant caractérisé par les 3 composantes couleurs  $R$ ,  $G$ , et  $B$ ).

La fonction de densité de chaque mode gaussien  $f_g(x; \mu, \sigma)$  est normalisée par le facteur adéquat (hauteur fixée à 1); cela signifie que le facteur  $\frac{1}{\sigma\sqrt{2\pi}}$  est ignoré, pour simplification ZENG et collab. [2008].

Ensuite, à partir de cette fonction d'appartenance de type-1, les auteurs proposent deux manières de composer un *IVFS* :

- L'imprécision peut d'abord être introduite au niveau du centre de la gaussienne. On considère alors que la véritable valeur  $\mu^*$  n'est pas précisément connue, mais qu'elle est comprise dans un intervalle de demi-longueur  $\Delta$  centré sur  $\mu$  :

$$\mu^* \in [\underline{\mu} = \mu - \Delta, \bar{\mu} = \mu + \Delta]. \quad (2.27)$$

La taille de l'intervalle d'incertitude sur la moyenne  $\Delta$  est définie en fonction de l'écart-type du mode gaussien, à l'aide d'un facteur  $k_m \geq 0$  :  $\Delta = k_m * \sigma$ .

L'empreinte d'incertitude (la *FOU*) de l'*IVFS* est alors définie comme la région contenant toutes les valeurs d'appartenance parcourues par la fonction d'appartenance  $f_g$ , en déplaçant son paramètre  $\mu^*$  dans cet intervalle d'incertitude. Le résultat est visible sur la figure 2.8.

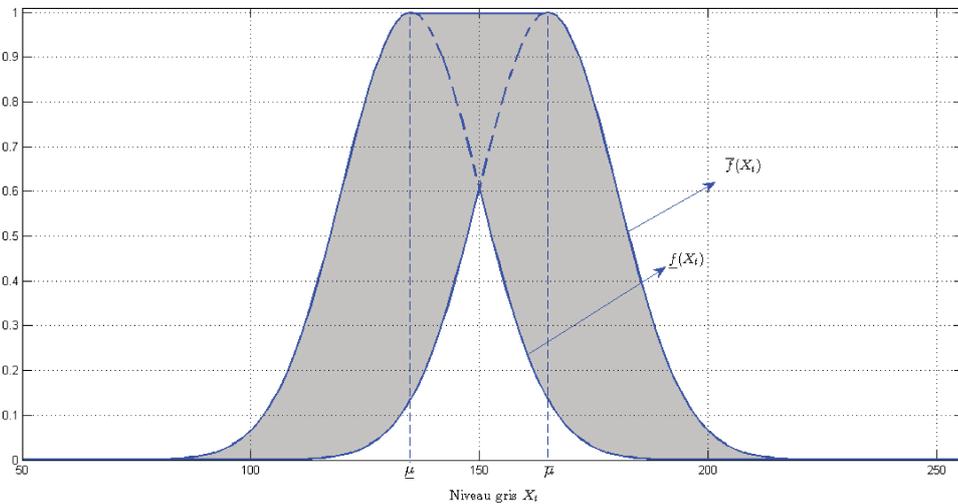


FIGURE 2.8 – *IVFS* d'un mode gaussien (incertitude sur la moyenne).

Les borne supérieure et inférieure de la *FOU* sont alors définies par :

$$\bar{f}(x; \mu, \sigma, \Delta) = \max_{\mu^* \in [\mu - \Delta, \mu + \Delta]} f_g(x; \mu^*, \sigma), \quad (2.28)$$

$$\underline{f}(x; \mu, \sigma, \Delta) = \min_{\mu \in [\mu^* - \Delta, \mu^* + \Delta]} f_g(x; \mu^*, \sigma). \quad (2.29)$$

- L'autre possibilité considérée par les auteurs consiste à intégrer l'incertitude au niveau du paramètre de dispersion de la gaussienne.

Dans ce cas, les bornes de la *FOU* sont définies ainsi :

$$\overline{f}(x; \mu, \sigma) = f_g(x; \mu, \overline{\sigma}), \quad (2.30)$$

$$\underline{f}(x; \mu, \sigma) = f_g(x; \mu, \underline{\sigma}), \quad (2.31)$$

où les bornes d'écart-types sont définies à l'aide du facteur  $0 < k_v < 1$  :

$$\underline{\sigma} = k_v \sigma \quad , \quad \overline{\sigma} = \frac{1}{k_v} \sigma. \quad (2.32)$$

La *FOU* obtenue est représentée sur la 2.9.

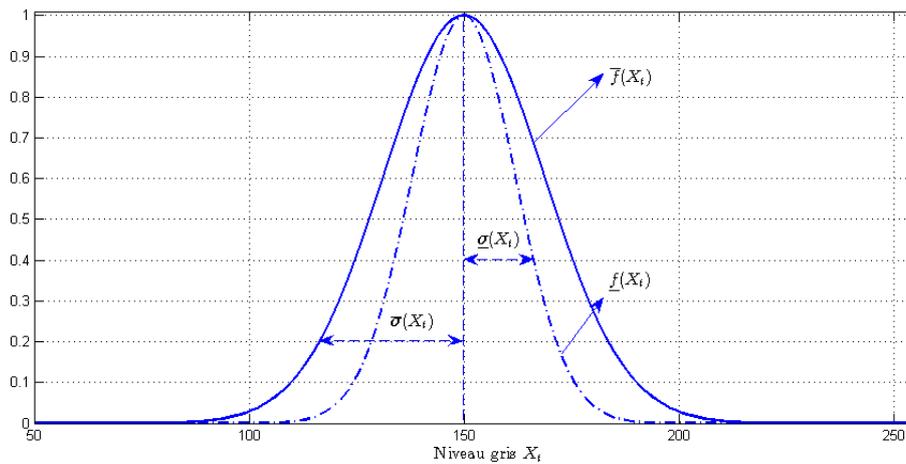


FIGURE 2.9 – *IVFS* d'un mode gaussien (incertitude sur l'écart-type).

Pour régler ces paramètres, qui déterminent directement la quantité d'imprécision de l'*IVFS* (soit sa *FOU*) Zeng *et al.* ont considéré qu'une gaussienne normalisée unidimensionnelle a 99,7% de sa masse probabiliste dans  $[\mu - 3\sigma, \mu + 3\sigma]$ . Ils ont donc proposé :  $k_m \in [0,3]$  et  $k_v \in [0.3,1]$  ZENG et collab. [2008] ZENG et LIU [2006a].

La figure 2.10 illustre comment on associe le mode gaussien flou à partir du modèle *GMM* classique stocké en mémoire, dans le cas où l'incertitude est insérée au niveau du centre du mode.

### Algorithme de la méthode

La méthode floue *T2-FGMM* est structurellement semblable à la technique nette proposée par STAUFFER et GRIMSON [1999] (cf. figure 1.7). Le passage à un modèle *GMM* imprécis affecte surtout l'étape **Étiquetage des pixels** (cf. section 2.4.3).

L'organigramme de la méthode est représenté sur la figure 2.11.

Voici maintenant la description de ses différentes étapes. Elles doivent être effectuées pour chacun des pixels (il y a en effet un modèle *GMM* par pixel).

Notons que les pixels peuvent être traités de manière indépendante, ce qui autorise une implémentation parallèle.

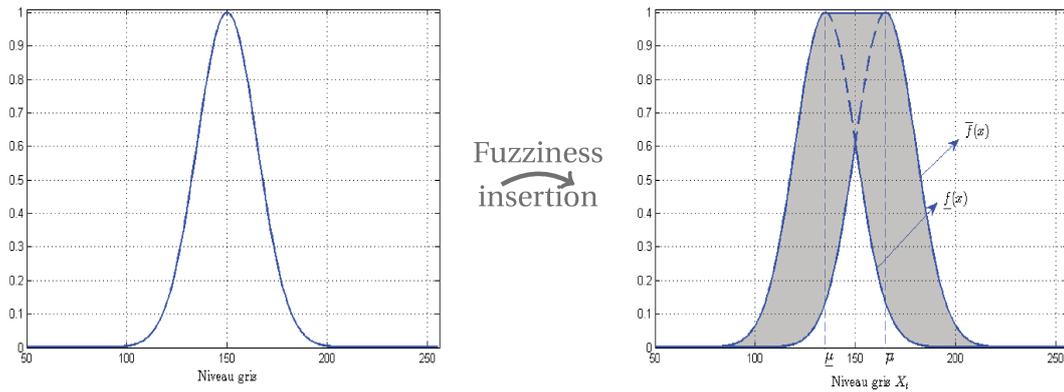


FIGURE 2.10 – Insertion du flou dans un mode gaussien.

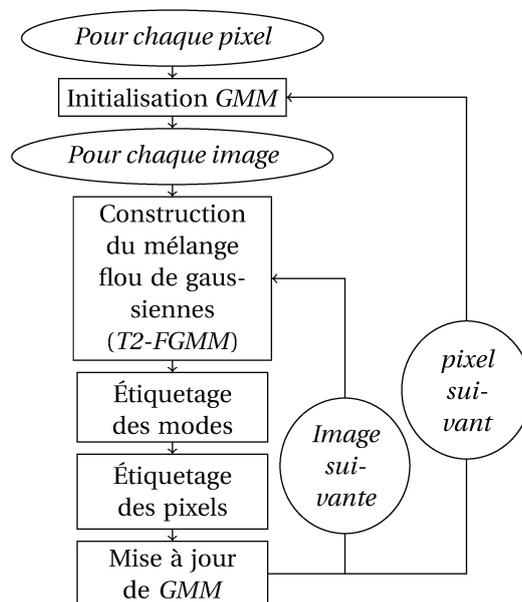


FIGURE 2.11 – Algorithme de la méthode T2-FGMM.

Les images utilisées dans les démonstrations de cette partie, proviennent d'une vidéo « maison », la vidéo *UL*, tournée à l'Université Libanaise : c'est une vidéo de dimensions  $640 \times 480$ , avec 29 images par seconde, et 723 images au total. Son intérêt principal réside dans le changement de lumière, ainsi que les mouvements de piétons.

### Étape 1 : initialisation du GMM

Les paramètres des modes gaussiens nets (ceux stockés en mémoire, qui n'intègrent pas encore l'incertitude floue) peuvent être initialisés à l'aide de l'algorithme espérance-maximisation (*EM*). Pour ce faire, il suffit d'extraire l'historique des niveaux du pixel sur une séquence d'images (voire sur toute la séquence, si elle est suffisamment petite, l'algorithme *EM* étant relativement lent).

La figure 2.12 montre un pixel particulier dans deux images différentes. En raison d'un personnage en mouvement, sa classe change entre ces deux images.

La figure 2.13 montre l'évolution de ce pixel dans le temps (en noir) et l'état de son GMM à l'image 100

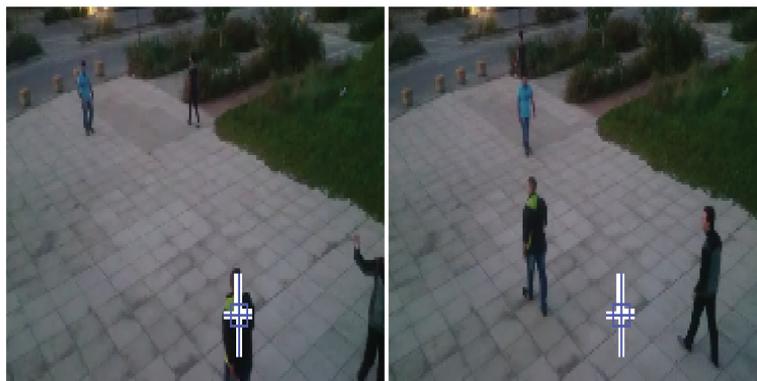


FIGURE 2.12 – Images 1 et 100 d'une séquence vidéo, avec marquage du pixel étudié.

(en rouge). Le creux important (en noir) dans les premières images est dû au personnage en mouvement. Il en résulte un mode gaussien *faiblement pondéré*, qui devrait normalement être interprété comme un mode de premier plan.

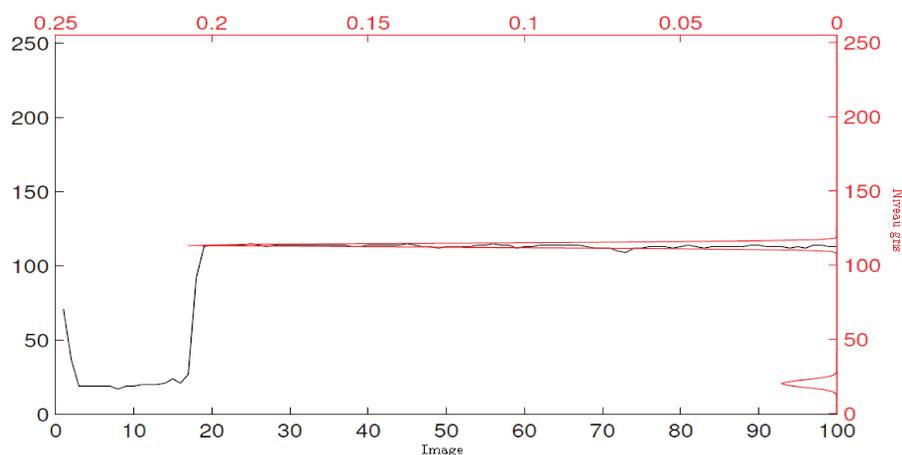


FIGURE 2.13 – Évolution de l'intensité du pixel sur les 100 premières images (en noir) et état du *GMM* à l'image 100 (en rouge).

Sur cette figure on peut voir qu'il y a des parties de la courbe noire qui ne sont pas couvertes par la courbe rouge, problème qui doit être pris en compte par le mélange de gaussiennes flou de *IVFS*.

## Étape 2 : Étiquetage des modes

Chaque mode gaussien est marqué soit *Fond* soit *Objet*. Cette étape est la même que dans le *GMM* classique (cf. section 1.6.3).

## Étape 3 : Étiquetage du pixel

La plupart des méthodes estiment la classe du pixel d'après la distance entre son niveau et le centre du mode le plus proche.

Mais c'est à cette étape que l'on introduit le flou : on dispose maintenant de deux fonctions d'appartenance  $\underline{f}$  et  $\overline{f}$ . Par construction, elles restent néanmoins centrées sur le centre  $\mu$  de la fonction nette  $f$ .

Bouwmans *et al.* BOUWMANS et EL BAF [2010]; EL BAF et collab. [2008b, 2009] basent leur décision sur la longueur de l'intervalle de « log-vraisemblance » au niveau  $x$  du pixel courant, soit :

$$L(x) = \left| \ln(\bar{f}(x)) - \ln(\underline{f}(x)) \right|. \quad (2.33)$$

Cet indicateur a d'abord été proposé par Zeng et Liu DUDA et collab. [2001] comme une mesure d'incertitude (une « part » d'entropie, cf. section 2.3.4) destinée à la détection d'*outliers*.

Cette quantité d'incertitude est fonction du facteur  $k_m \geq 0$ , et on peut montrer qu'elle vaut :

— Dans le cas d'une incertitude introduite sur la moyenne,

$$L(x) = \begin{cases} 2k_m|x - \mu|/\sigma, & \text{si } x \leq \underline{\mu} \quad \text{ou} \quad x \geq \bar{\mu} \\ |x - \mu|^2/2\sigma^2 + k_m|x - \mu|/\sigma + k_m^2/2, & \text{si } \underline{\mu} < x < \bar{\mu}. \end{cases} \quad (2.34)$$

— Dans le cas d'une incertitude sur l'écart-type :

$$L(x) = \frac{(\frac{1}{k_v^2} - k_v^2)|x - \mu|}{2\sigma}. \quad (2.35)$$

Dans BOUWMANS et EL BAF [2010]; EL BAF et collab. [2008b, 2009], la condition d'affectation du pixel, de niveau  $x$ , à l'un des  $K$  modes gaussiens est la suivante :

$$L(x) < k_p\sigma, \quad (2.36)$$

$k_p$  étant un coefficient constant à ajuster suivant la vidéo traitée.

D'après les équations qui précèdent, on observe que  $L(x)$  est fonction croissante de  $|x - \mu|/\sigma$  : on retrouve donc la règle usuelle, qui apparaît notamment dans la méthode de Bayes : un pixel est associé à un mode, s'il est suffisamment vraisemblable, et donc suffisamment proche de son centre (relativement à son paramètre d'écart-type).

Cette condition est évaluée pour l'ensemble des modes gaussiens, dans leur ordre de priorité  $\frac{w}{\sigma}$  (cf. méthode de Stauffer dans la section 1.6.3). Le mode qui l'emporte est le premier à passer le test : le plus prioritaire.

Deux cas peuvent alors se présenter :

- Une correspondance est trouvée avec l'un des  $K$  modes gaussiennes. Dans ce cas, la distribution gaussienne est identifiée comme *Fond* s'il appartient aux  $K_B$  premiers modes, comme *Objet* sinon.
- Aucune correspondance. Dans ce cas, le pixel est classé comme *Objet*.

On rentre alors la décision dans un masque binaire de la taille de l'image.

#### Étape 4 : Mise à jour du GMM

En préalable à la classification du pixel de la prochaine image, les paramètres de son GMM doivent être mis à jour.

La mise à jour s'effectue suivant les règles 1.14 de la méthode de Stauffer 1.6.3.

Une fois la mise à jour du modèle effectuée - sur tous les pixels -, la détection des objets sur l'image suivante peut débuter, en reprenant à l'étape 2.

#### 2.4.4 Illustration de la méthode *T2-FGMM*

Les expérimentations réalisées dans cette partie l'ont été grâce à une implémentation sous MATLAB™ d'une version de la méthode *T2-FGMM-UM* dédiée aux vidéos en niveaux de gris (pour simplifier la représentation des modes gaussiens).

Elles ont été obtenues en traitant la vidéo *UL* (de l'Université Libanaise) présentée en section 2.4.3.

Nous avons fixé les paramètres de la méthode aux valeurs suivantes, choisies pour leurs bonnes performances :

- $K = 3$ , le nombre de modes gaussiens;
- $k_m = 1,5$ , le coefficient qui contrôle l'incertitude du centre des modes;
- $T_b = 0,75$ , le seuil qui sert à déterminer le nombre de modes *Fond*  $K_B$ ;
- $k_p = 1,5$ , le coefficient qui règle le seuil d'affectation à un mode.

La figure 2.14 montre les images 1 et 500 de la vidéo *UL*. Sur l'image, on a indiqué la localisation du pixel étudié. On peut constater l'importance de sa différence d'éclairage (raison pour laquelle il a été choisi).



FIGURE 2.14 – Deux images de la vidéo *UL*, avant et après le changement d'éclairage.

La figure 2.15 représente l'état du *GMM* de ce pixel à l'image 1 (initialisé par *EM*), avant le changement de luminosité. Les modes gaussiens étiquetés *Objet* sont dessinés en rouge, les modes *Fond* en noir. La barre verticale montre l'intensité du pixel : elle est également en noir car dans cette image 1, le pixel est classé *Fond*; il le sera en fait systématiquement sur l'ensemble de la vidéo.

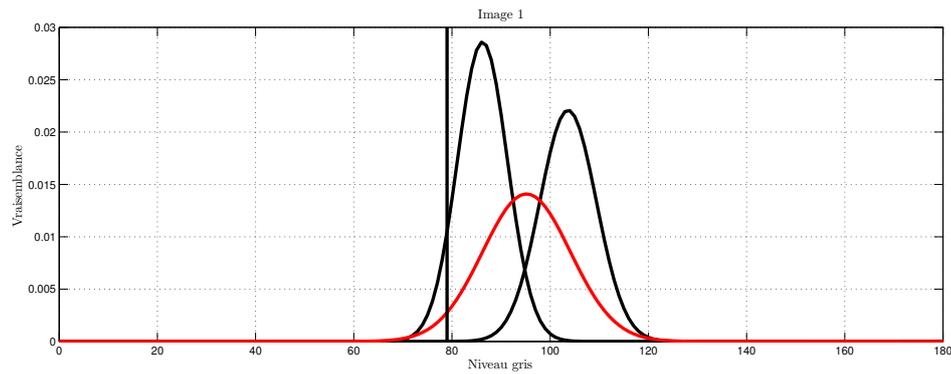


FIGURE 2.15 – *GMM* net à l'image 1, relatif au pixel marqué sur l'image précédente.

La figure 2.16 illustre l'état des  $K = 3$  modes gaussiens, dans l'étape d'étiquetage des pixels : le flou a été introduit, ce sont les *IVFS* obtenus à partir des modes gaussiens nets. Ils sont dessinés dans l'ordre de priorité (définis par l'indicateur  $\frac{w_k}{\sigma_k}$ ), avec le même code couleur. La barre du niveau de gris du pixel est verte, là où l'affectation a réussi, noire sinon. Le pixel est associé au premier mode, et est bien classé *Fond*.

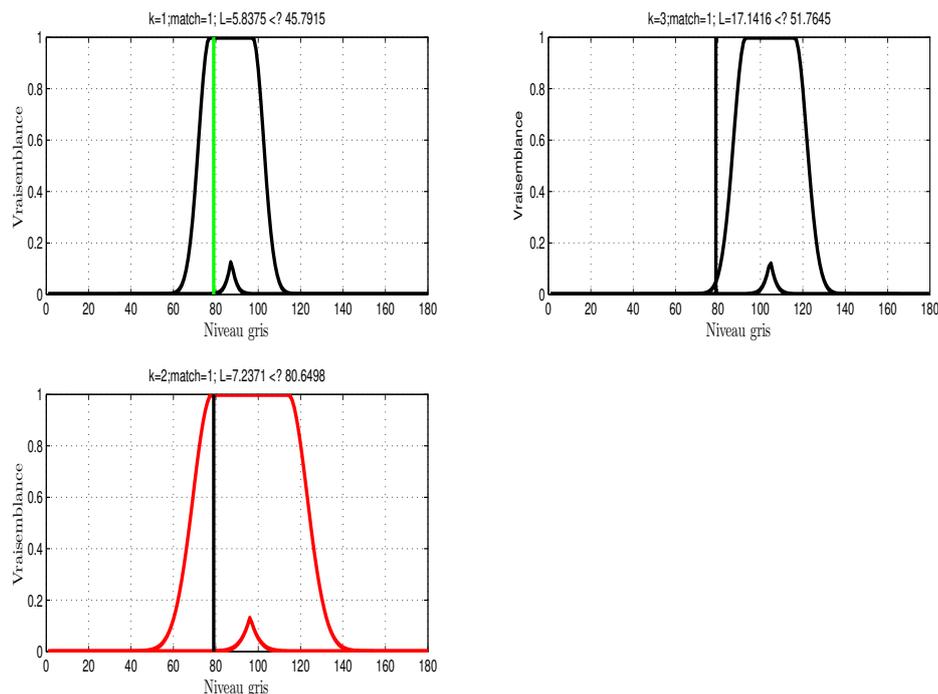


FIGURE 2.16 – *GMM IVFS* à l'image 1, relatif au pixel étudié.

Entre l'image 1 et l'image 500, la luminosité globale de l'image a fortement augmenté. On observe sur l'état des gaussiennes à l'image 500 (cf. figure 2.17), que le pixel est encore classé *Fond*. Les autres modes n'apparaissent pas, car de poids quasi-nuls. Ce qui est visible sur la représentations des modes *IVFS* sur la figure 2.18.

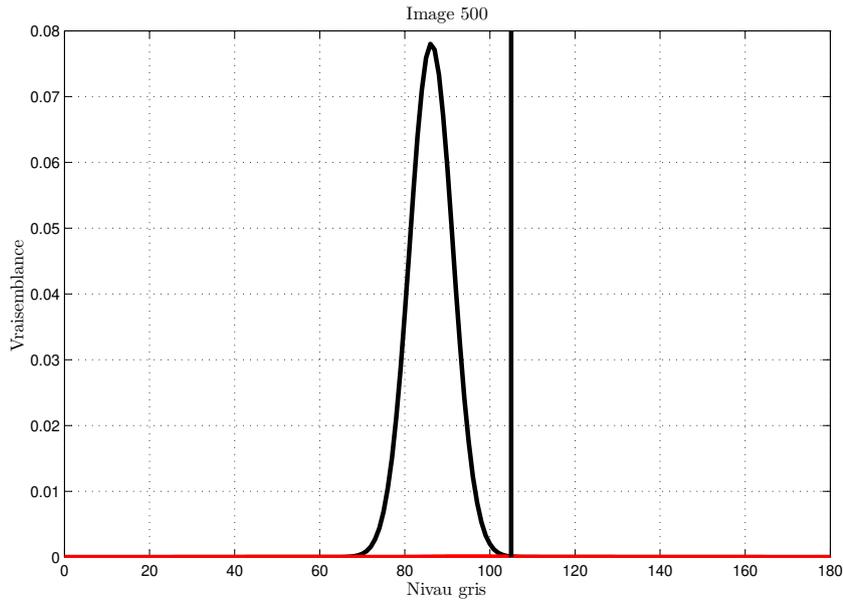


FIGURE 2.17 – *GMM* net à l'image 500, relatif au pixel étudié.

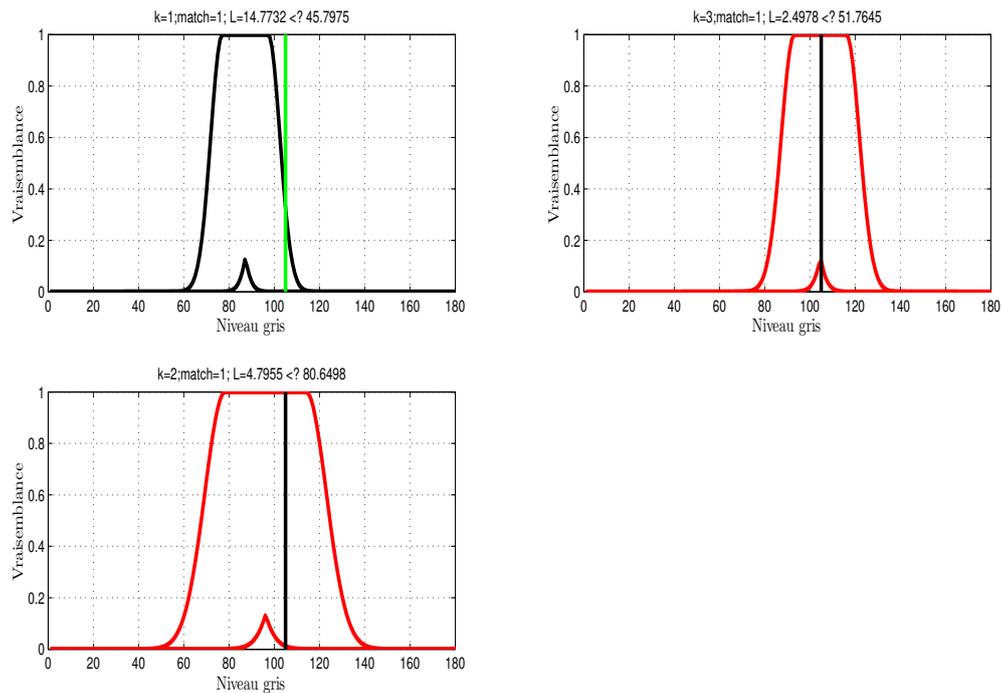


FIGURE 2.18 – *GMM IVFS* à l'image 500, relatif au pixel étudié.

### 2.4.5 Critiques de la méthode *T2-FGMM*

Cette méthode a été proposée pour modéliser l'incertitude en utilisant le flou de type-2, mais nous pouvons souligner quelques points faibles :

- Le flou inséré sous forme d'*IVFS* n'intervient en définitive que dans l'étape de classification du pixel, plus précisément dans le calcul de l'écart entre les deux log-vraisemblances  $L(x)$ . Or cet écart se

résumé - comme on a pu le voir - à une distance entre le centre du mode et le niveau du pixel, ou encore à la vraisemblance du pixel suivant la gaussienne nette. Le flou n'est donc pas réellement nécessaire.

- Ensuite, toujours dans cette même étape, l'écart de vraisemblance est aussitôt seuillé. On ne calcule pas de niveaux d'appartenance floue aux décisions *Fond* et *Objet*. Or cela pourrait être utile à des fusions avec des décisions du même type provenant d'autres informations (texture par exemple).
- Enfin, si deux modes gaussiens sont très proches l'un de l'autre et que le pixel est associé à l'un d'eux - le plus prioritaire -, l'autre mode n'aura aucun impact sur la décision : il pourrait être intéressant de conserver trace de cette ambiguïté, toujours dans l'optique de fusions ultérieures.

Les deux dernières critiques sont très secondaires. Néanmoins, l'agrégation de plusieurs décisions semble très en vogue, et on cherchera donc - dans la méthode proposée dans le prochain chapitre - à en tenir compte.

## 2.5 Conclusion du chapitre

Les différents ensembles flous (type-1, type-2, intuitifs, etc.) constituent un cadre formel intéressant pour la modélisation des incertitudes.

Dans ce chapitre nous avons discuté de l'intégration de modèles flous dans les techniques de soustraction de fond. Nous avons présenté en particulier les nombres flous de type *IVFS* et leur intérêt en détection d'arrière-plan. Ils permettent la prise en compte des imperfections associées à la vidéo de façon élégante, rigoureuse et raisonnablement complexe.

Enfin une méthode de soustraction basée sur un modèle de mélange gaussiennes floues de type *IVFS*, choisie pour sa simplicité et son modèle *GMM* flou, a été détaillée. Elle a ensuite fait l'objet d'expérimentations à des fins d'illustration, mais aussi de quelques critiques.

Nous allons présenter au chapitre suivant la méthode de modélisation d'arrière-plan de vidéo que nous proposons. Il s'agit d'une nouvelle méthode qui utilise le même modèle *GMM* flou de type *IVFS*, mais avec une utilisation plus poussée du flou. En particulier, dans l'optique de profiter un maximum de l'information floue introduite dans le modèle considéré <https://sharelatex.univ-littoral.fr/project/59f6e4cdddb2830b0f31ac8b>, la décision finale du pixel est élaborée par fusion de plusieurs décisions intermédiaires. Ces dernières sont exploitées pour minimiser les erreurs de classification résiduelles, et elles proviennent de la prise en compte de la continuité spatiale et temporelle du statut *Fond/Objet* des pixels.

## 2.6 Références

ATANASSOV, K. et G. GARGOV. 1989, «Interval valued intuitionistic fuzzy sets», *Fuzzy sets and systems*, vol. 31, n° 3, p. 343–349. [36](#)

ATANASSOV, K. T. 1986, «Intuitionistic fuzzy sets», *Fuzzy sets and Systems*, vol. 20, n° 1, p. 87–96. [36](#)

ATANASSOV, K. T. 1999, «Intuitionistic fuzzy sets», dans *Intuitionistic Fuzzy Sets*, Springer, p. 1–137. [36](#), [38](#)

- BIGAND, A. et O. COLOT. 2010, «Fuzzy filter based on interval-valued fuzzy sets for image filtering», *Fuzzy Sets and Systems*, vol. 161, n° 1, p. 96–117. [43](#)
- BLOCKLEY, D. I. et P. GODFREY. 2000, *Doing it differently : systems for rethinking construction*, Thomas Telford. [32](#)
- BOUWMANS, T. 2012, «Background subtraction for visual surveillance : A fuzzy approach», *Handbook on soft computing for video surveillance*, vol. 5, p. 103–138. [44](#), [45](#)
- BOUWMANS, T. et F. EL BAF. 2010, «Modeling of dynamic backgrounds by type-2 fuzzy gaussian mixture models», *MASAUM Journal of Basic and Applied Sciences*, vol. 1, n° 2, p. 265–276. [44](#), [46](#), [51](#)
- BURILLO, P. et H. BUSTINCE. 1996, «Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets», *Fuzzy sets and systems*, vol. 78, n° 3, p. 305–316. [42](#)
- CHAIRA, T. et A. RAY. 2008, «A new measure using intuitionistic fuzzy set theory and its application to edge detection», *Applied Soft Computing*, vol. 8, n° 2, p. 919–927. [38](#)
- CHIRANJEEVI, P. et S. SENGUPTA. 2014, «Neighborhood supported model level fuzzy aggregation for moving object segmentation», *IEEE Transactions on Image Processing*, vol. 23, n° 2, p. 645–657. [45](#)
- DE LUCA, A. et S. TERMINI. 1972, «A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory», *Information and control*, vol. 20, n° 4, p. 301–312. [42](#)
- DELEPOULLE, S., A. BIGAND et C. RENAUD. 2011, «Interval type-2 fuzzy sets-based no-reference quality evaluation of synthetic images», *Proceedings of CGVR*, vol. 11. [40](#), [43](#)
- DESCRIJVER, G. et E. E. KERRE. 2007, «On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision», *Information Sciences*, vol. 177, n° 8, p. 1860–1866. [38](#)
- DESPI, I., D. OPRIS et E. YALCIN. 2013, «Generalised atanassov intuitionistic fuzzy sets», dans *Proceeding of the Fifth International Conference on Information, Process, and Knowledge Management*. [37](#)
- DUDA, R. O., P. E. HART et D. G. STORK. 2001, «Pattern classification second edition john wiley & sons», *New York*, vol. 58. [51](#)
- EL BAF, F., T. BOUWMANS et B. VACHON. 2008a, «Fuzzy integral for moving object detection», dans *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, IEEE, p. 1729–1736. [44](#)
- EL BAF, F., T. BOUWMANS et B. VACHON. 2008b, «Type-2 fuzzy mixture of gaussian model : application to background modeling», dans *International Symposium on Visual Computing*, Springer, p. 772–781. [44](#), [46](#), [51](#)
- EL BAF, F., T. BOUWMANS et B. VACHON. 2009, «Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos», dans *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, IEEE, p. 60–65. [44](#), [46](#), [51](#)

## CHAPITRE 2. ÉTAT DE L'ART SUR L'APPLICATION DES ENSEMBLES FLOUS À LA SOUSTRACTION DU FOND EN VIDÉO

---

- ENV13005, N. 1999, «Guide pour l'expression de l'incertitude de mesure», Paris, France : AFNOR. 32
- GUTTI, V. et C. SHANKAR. 2014, «A novel approach to background subtraction using fuzzy color histogram», *Journal of Advanced Engineering Technologies*, vol. 3, n° 3, p. 231–239. 45
- HISDAL, E. 1981, «The if then else statement and interval-valued fuzzy sets of higher type», *International Journal of Man-Machine Studies*, vol. 15, n° 4, p. 385–455. 38
- HOSSEINI, R., J. DEHMESHKI, S. BARMAN, M. MAZINANI et S. QANADLI. 2010, «A genetic type-2 fuzzy logic system for pattern recognition in computer aided detection systems», dans *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, IEEE, p. 1–7. 40
- HWANG, C. et F. C.-H. RHEE. 2007, «Uncertain fuzzy clustering : interval type- 2 fuzzy approach to c-means», *Fuzzy Systems, IEEE Transactions on*, vol. 15, n° 1, p. 107–120. 39
- IBWM, S., G. IEC et AL. 1995, *Guide pour l'expression de l'incertitude de mesure*, ISO, Geneva. URL <https://cds.cern.ch/record/641137>. 31
- KARNIK, N. N., J. M. MENDEL et Q. LIANG. 1999, «Type-2 fuzzy logic systems», *Fuzzy Systems, IEEE Transactions on*, vol. 7, n° 6, p. 643–658. 40
- KAUFMANN, A. 1975, *Introduction to the theory of fuzzy subsets*, vol. 2, Academic Pr. 42
- KIM, W. et C. KIM. 2012, «Background subtraction for dynamic texture scenes using fuzzy color histograms», *IEEE Signal processing letters*, vol. 19, n° 3, p. 127–130. 45
- LU, X., T. IZUMI, T. TAKAHASHI et L. WANG. 2014, «Moving vehicle detection based on fuzzy background subtraction», dans *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, IEEE, p. 529–532. 44
- MANJULA, D. et M. SIVABALAKRISHNAN. 2010, «Adaptive background subtraction in dynamic environments using fuzzy logic», *International Journal of Video & Image Processing and Network Security*, vol. 10, p. 13–16. 45
- PAL, N. et J. BEZDEK. 1994, «Measures of fuzziness : a review and several classes», . 42
- R.BOUKEZZOULA, S. et L.FOULLOY. 2016, «Sur les systèmes flous de type-2 en contrôle!», dans *LFA 2016*. 39
- ROSELL-ORTEGA, J., G. GARCIA-ANDREU, A. RODAS-JORDA et V. ATIENZA-VANACLOIG. 2010, «A combined self-configuring method for object tracking in colour video», dans *Pattern Recognition (ICPR), 2010 20th International Conference on*, IEEE, p. 2081–2084. 44
- SAMBUC, R. 1975, *Fonctions and Floues : Application a l'aide au Diagnostic en Pathologie Thyroïdienne*, Faculté de Médecine de Marseille. 43

- SHRESTHA, R. R., A. BÁRDOSSY et F. NESTMANN. 2007, «Analysis and propagation of uncertainties due to the stage–discharge relationship : a fuzzy set approach», *Hydrological Sciences Journal*, vol. 52, n° 4, p. 595–610. [35](#)
- SIGARI, M. 2008, «Fuzzy background modeling/subtraction and its application in vehicle detection», dans *World Congress on Engineering and Computer Science, WCECS 2008*. [44](#)
- SIGARI, M. H., N. MOZAYANI et H. POURREZA. 2008, «Fuzzy running average and fuzzy background subtraction : concepts and application», *International Journal of Computer Science and Network Security*, vol. 8, n° 2, p. 138–143. [44](#)
- STAUFFER, C. et W. E. L. GRIMSON. 1999, «Adaptive background mixture models for real-time tracking», dans *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, IEEE. [46, 48](#)
- SUBUDHI, B. N., S. GHOSH, S.-B. CHO et A. GHOSH. 2016, «Integration of fuzzy markov random field and local information for separation of moving objects and shadows», *Information Sciences*, vol. 331, doi :<https://doi.org/10.1016/j.ins.2015.10.031>, p. 15 – 31, ISSN 0020-0255. URL <http://www.sciencedirect.com/science/article/pii/S0020025515007598>. [44](#)
- SUGENO. 2013, «Categories of uncertainty», dans *Exploring structure of Uncertainty, Symposium on*, LAMIH, Valenciennes, p. 1–41. [31](#)
- TIZHOOSH, H. R. 2005, «Image thresholding using type ii fuzzy sets», *Pattern recognition*, vol. 38, n° 12, p. 2363–2372. [43](#)
- WANG, D. et L. ACAR. 1999, «An analysis of type-1 and type-2 fuzzy logic systems», dans *Intelligent Control/Intelligent Systems and Semiotics, 1999. Proceedings of the 1999 IEEE International Symposium on*, IEEE, p. 353–358. [43](#)
- WIERMAN, M. J. 2010, «An introduction to the mathematics of uncertainty», *Creighton University*. [35](#)
- WREN, C. R., A. AZARBAYEJANI, T. DARRELL et A. P. PENTLAND. 1997, «Pfinder : Real-time tracking of the human body», *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, n° 7, p. 780–785. [44](#)
- YAGER, R. R. 1979, «On the measure of fuzziness and negation part i : membership in the unit interval», . [42](#)
- YAGER, R. R. 2013, «Pythagorean fuzzy subsets», dans *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint*, IEEE, p. 57–61. [38](#)
- YAGER, R. R. 2014, «Pythagorean membership grades in multicriteria decision making», *IEEE Transactions on Fuzzy Systems*, vol. 22, n° 4, p. 958–965. [38](#)
- ZADEH, L. A. 1965, «Fuzzy sets», *Information and control*, vol. 8, n° 3, p. 338–353. [33](#)

- ZENG, J. et Z.-Q. LIU. 2006a, «Type-2 fuzzy hidden markov models and their application to speech recognition», *IEEE Transactions on Fuzzy Systems*, vol. 14, n° 3, p. 454–467. [48](#)
- ZENG, J. et Z.-Q. LIU. 2006b, «Type-2 fuzzy sets for handling uncertainty in pattern recognition», dans *Fuzzy Systems, 2006 IEEE International Conference on*, IEEE, p. 1247–1252. [46](#)
- ZENG, J. et Z.-Q. LIU. 2007, «Type-2 fuzzy sets for pattern recognition : The state-of-the-art», *Journal of Uncertain Systems*, vol. 1, n° 3, p. 163–177. [46](#)
- ZENG, J., L. XIE et Z.-Q. LIU. 2008, «Type-2 fuzzy gaussian mixture models», *Pattern Recognition*, vol. 41, n° 12, p. 3636–3643. [46](#), [47](#), [48](#)
- ZHANG, H. et D. XU. 2006, «Fusing color and texture features for background model», *Fuzzy Systems and Knowledge Discovery*, p. 887–893. [44](#)
- ZI-LONG, L., L. WEI-MING et Z. YANG. 2013, «A background modeling method based on adaptive fuzzy estimation», *Journal of South China University of Technology (Natural Science Edition)*, vol. 9, p. 014. [45](#)



# Chapitre 3

## La méthode *IV-FGMM* : soustraction du fond basée sur un mélange de gaussiennes imprécises

### Sommaire

---

<b>3.1 Introduction</b> . . . . .	<b>63</b>
<b>3.2 Algorithme proposé</b> . . . . .	<b>63</b>
3.2.1 Organigramme . . . . .	63
3.2.2 Étapes principales . . . . .	64
<b>3.3 Décision floue</b> . . . . .	<b>65</b>
3.3.1 Agrégation des intervalles imprécis . . . . .	65
3.3.2 Construction de la décision floue par comparaison d'intervalles . . . . .	67
3.3.3 Décision nette . . . . .	68
<b>3.4 Décisions complémentaires</b> . . . . .	<b>68</b>
3.4.1 Décision spatiale floue . . . . .	68
3.4.2 Décision temporelle floue . . . . .	70
<b>3.5 Fusion des décisions</b> . . . . .	<b>70</b>
3.5.1 Décision finale floue . . . . .	70
3.5.2 Défuzzification de la décision finale floue . . . . .	71
<b>3.6 Autres étapes</b> . . . . .	<b>71</b>
3.6.1 Mise à jour du modèle <i>GMM</i> . . . . .	71
3.6.2 Fusion des modes proches . . . . .	72

**3.7 Conclusion du chapitre** . . . . . 72

**3.8 Références** . . . . . 72

---

## 3.1 Introduction

Dans le chapitre précédent nous avons cité plusieurs méthodes de soustraction de l'arrière-plan. Nous nous sommes plus particulièrement intéressés à la méthode *T2-FGMM* [BOUWMANS et EL BAF \[2009\]](#), pour sa gestion de l'incertitude basée sur des sous-ensembles flous de type-2.

Ce chapitre présente une méthode voisine de *T2-FGMM*, dont l'originalité réside dans son processus de décision. Le flou y est conservé jusqu'à la décision binaire finale. Et l'incertitude est exploitée pour pondérer la participation de nouvelles « sources » d'information au processus de décision.

Ces informations supplémentaires proviennent d'une hypothèse d'homogénéité spatiale et temporelle de la classe d'un pixel dans une séquence vidéo. On peut en effet supposer que la classe d'un pixel au sein d'une image, *Fond* ou *Objet*, ne dépende pas seulement de son niveau propre dans cette image, mais aussi d'un contexte spatial et temporel.

Dans la méthode proposée, la classe d'un pixel d'une image peut être fonction de celle de ses voisins (contexte spatial), ainsi que celle de son ou ses prédécesseur(s) dans les images antérieures (contexte temporel).

Et ce contexte sera d'autant plus utile dans les situations où la décision classique est difficile. Il arrive en effet que les méthodes basées sur un *GMM* ne parviennent pas à classer un pixel : à cause de la proximité des modes gaussiens par exemple (ambiguïté du modèle), ou d'un niveau de pixel hors-norme (rejet en distance).

Dans de tels cas, il est intéressant de quantifier l'incertitude de cette décision élémentaire, pour évaluer dans quelle mesure elle doit être enrichie par l'information contextuelle évoquée ci-dessus.

conducteurs, etc. Mais la tâche n'est pas toujours aisée. En effet, les vidéos réelles mettent souvent en défaut les algorithmes de soustraction de fond, à cause de perturbations diverses. Parmi elles, les arrière-plans dynamiques (dus à des changements de luminosité ou des mouvements) sont particulièrement fréquents et gênants.

## 3.2 Algorithme proposé

### 3.2.1 Organigramme

La méthode proposée est structurellement assez proche de la méthode *GMM* classique [STAUFFER et GRIMSON \[1999\]](#) et de la méthode floue *T2-FGMM* [BOUWMANS et EL BAF \[2009\]](#); [EL BAF et collab. \[2008\]](#).

Elle s'en distingue principalement par la manière dont elle exploite un modèle *GMM* flou de type-2, et par la fusion de différentes décisions, spatiale et/ou temporelle.

Nous l'avons appelée *IV-FGMM*, pour *Interval-Valued Fuzzy Gaussian Mixture Model*, en référence à la nature du sous-ensemble de type-2 qu'elle utilise.

Son organigramme est présenté sur la figure [3.1](#).

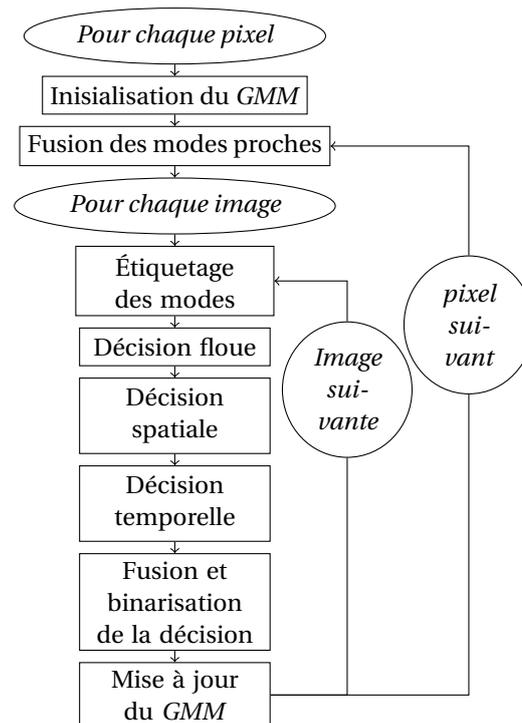


FIGURE 3.1 – Algorithme de la méthode proposée.

### 3.2.2 Étapes principales

L'algorithme de notre méthode s'articule autour des étapes suivantes, les mêmes que celles de la technique originelle [STAUFFER et GRIMSON \[1999\]](#) :

- **Initialisation du GMM** : le modèle *GMM* de chaque pixel est initialisé en fonction d'un historique de ses niveaux (ou sur une partie de la vidéo). À ce stade, il s'agit d'un modèle *GMM* classique, probabiliste.

Son initialisation est réalisée par l'algorithme usuel *EM*.

L'étape itérée de mise à jour des modes étant capable d'adapter le modèle aux changements de la vidéo, cette étape n'est pas indispensable. Mais elle offre l'avantage de rendre les premières décisions plus pertinentes.

- **Fusion des modes proches** : *EM* étant susceptible de fournir des modes superposés (notamment parce qu'il est difficile de prévoir le nombre de modes), cette étape a pour but d'éliminer les modes trop ambigus.
- **Étiquetage des modes** : chaque mode gaussienne est marqué soit *Fond* soit *Objet*, suivant la procédure utilisée avec le *GMM* classique [STAUFFER et GRIMSON \[1999\]](#). On trie de la même façon les  $K$  gaussiennes par ordre de poids  $\frac{w_k}{\sigma_k}$  décroissants. Les  $K_B$  premiers modes gaussiennes sont étiquetés *Fond*, les autres *Objet*. La valeur de  $K_B$  est obtenue en déterminant le nombre minimal de premiers modes, dont le cumul des poids dépasse un seuil  $T_b \in [0,1]$  (cf. équation 1.12).
- **Décision floue** : après transformation du modèle *GMM* probabiliste en *GMM* flou d'ordre 2, une première décision floue est prise en calculant les intervalles d'appartenance du niveau du pixel à chacun des modes. La technique est inspirée de celle de Zeng 2.28, mais tous les modes sont mis à

contribution dans la décision.

Tous les calculs de décisions sont détaillés dans la section suivante.

- **Décision spatiale** : il s'agit ici de fusionner les décisions floues des pixels voisins.
- **Décision temporelle** : la continuité temporelle des pixels - malgré leur éventuel déplacement - est effectuée grâce à un calcul de flot optique.
- **Fusion des décisions** : les décisions spatiales et temporelles sont fusionnées, puis défuzzifiées, pour obtenir la classe finale du pixel dans l'image.
- **Mise à jour du GMM** : les paramètres du *GMM* sont mis à jour comme dans la méthode *T2-FGMM* et celle de Stauffer [STAUFFER et GRIMSON \[1999\]](#), d'après les équations [1.14](#).

## 3.3 Décision floue

### 3.3.1 Agrégation des intervalles imprécis

À l'inverse de la méthode *T2-FGMM* - également basée sur un modèle *GMM* flou -, nous proposons de construire une mesure floue de « préférence » entre les 2 décisions possibles (*Objet*, *Fond*), accompagnée d'un degré de « rejet », destiné à marquer l'absence de connaissance.

Pour cela, il s'agit d'abord d'associer au niveau  $x_t^p$  du pixel  $p$  considéré dans l'image  $t$ , un degré d'appartenance à chacun des modes gaussiens imprécis (d'indice  $k$ ), sous la forme d'un intervalle normalisé (*IVFS*) :

$$\left[ \underline{f}_k(x_t^p), \bar{f}_k(x_t^p) \right]. \quad (3.1)$$

La figure [3.2](#) illustre l'état d'un modèle *GMM* flou de type-2, composé de trois modes gaussiens. On observe leurs intervalles d'appartenance, en pointillés :  $F1_t(p), F2_t(p)$  sont deux intervalles *Objet* (en rouge),  $B1_t(p)$  est un intervalle *Fond* (en noir).

Ensuite, pour chaque classe de mode (*Objet*, *Fond*), les intervalles correspondants sont agrégés. La figure [3.3](#) montre les deux modes agrégés, ainsi que  $R$ , l'intervalle « seuil de rejet ».

L'agrégation normalisée est définie par une moyenne pondérée, dont les poids sont ceux proposés par Zeng pour trier les modes (par ordre décroissant de vraisemblance de leur classement comme *Fond*) :

$$w'_k = \rho \left( \frac{w_k}{\sigma_k} \right), \quad (3.2)$$

avec  $\rho(x) = x^2$ .

Les intervalles participant à la décision à l'issue de l'étape d'agrégation (cf. figure [3.3](#)) sont :

- $F_t(p)$  : l'intervalle de vraisemblance de l'hypothèse « valeur de pixel correspondant à une région *Objet* ».
- $B_t(p)$  : l'intervalle de vraisemblance de l'hypothèse *Fond*.

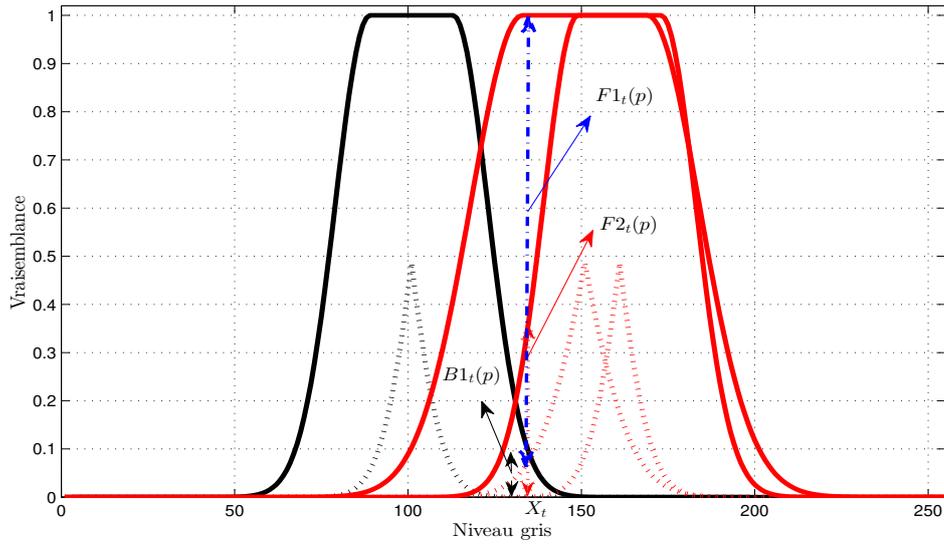


FIGURE 3.2 – Intervalles flous des modes gaussiens, avant agrégation.

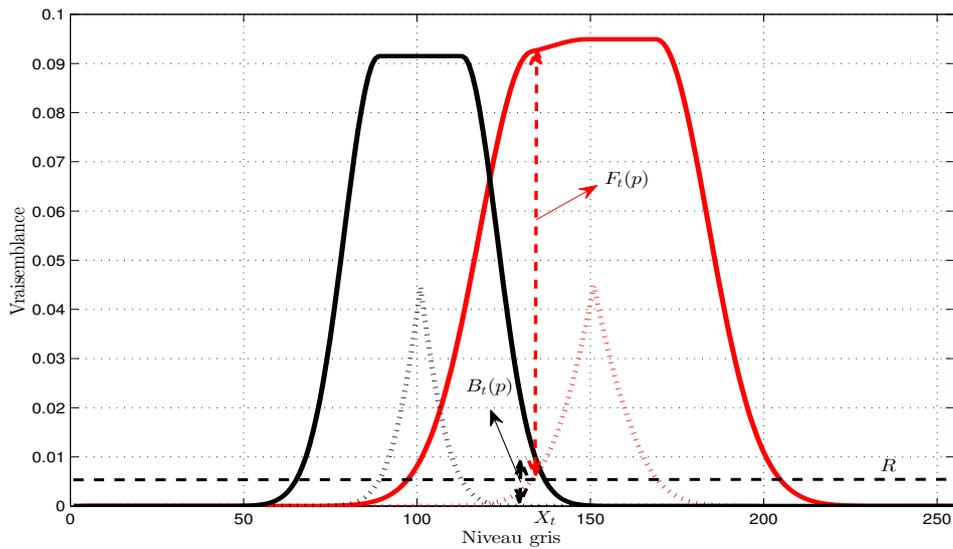


FIGURE 3.3 – Intervalles flous agrégés + Zone de rejet.

- $\mathbf{R} = [0, k_r]$  : l'intervalle seuil de *Rejet*; il représente les niveaux de vraisemblance non significatifs, définis a priori comme inférieurs à  $k_r \in \mathbb{R}^+$ .

Voici leurs formules, basées sur la moyenne pondérée :

$$\mathbf{F}_t(p) = \frac{1}{\sum_{k=1}^K w'_k} \left[ \sum_{k=K_B+1}^K w'_k f_{\underline{k}}(x_t^p), \sum_{k=K_B+1}^K w'_k \bar{f}_k(x_t^p) \right], \quad (3.3)$$

$$\mathbf{B}_t(p) = \frac{1}{\sum_{k=1}^K w'_k} \left[ \sum_{k=1}^{K_B} w'_k f_{\underline{k}}(x_t^p), \sum_{k=1}^{K_B} w'_k \bar{f}_k(x_t^p) \right], \quad (3.4)$$

$$\mathbf{R} = [0, k_r]. \quad (3.5)$$

### 3.3.2 Construction de la décision floue par comparaison d'intervalles

Nous proposons d'utiliser une mesure de comparaison d'intervalles, basée sur celle de [SENGUPTA et PAL \[2000\]](#), que nous décrivons ci-dessous.

Chaque intervalle  $\mathbf{A} = [a^-, a^+]$  est représenté par le couple de valeurs  $(c(\mathbf{A}), \delta(\mathbf{A}))$ , avec :

—  $c(\mathbf{A})$ , le centre de l'intervalle  $\mathbf{A}$  :

$$c(\mathbf{A}) = \frac{1}{2} (a^- + a^+), \quad (3.6)$$

—  $\delta(\mathbf{A})$ , la demi-largeur de  $\mathbf{A}$  :

$$\delta(\mathbf{A}) = \frac{1}{2} (a^+ - a^-). \quad (3.7)$$

L'indice d'acceptabilité mesure la vraisemblance de « l'hypothèse d'infériorité du premier intervalle par rapport au second » :

$$\mathcal{A}'_{\ominus}(\mathbf{A}, \mathbf{B}) = \frac{c(\mathbf{B}) - c(\mathbf{A})}{\delta(\mathbf{B}) + \delta(\mathbf{A})}, \text{ si } c(\mathbf{A}) \leq c(\mathbf{B}), \quad (3.8)$$

et si  $\delta(\mathbf{A}) + \delta(\mathbf{B}) \neq 0$  (mais les niveaux de vraisemblance étant issues de gaussiennes, on pourra faire systématiquement cette hypothèse).

On remarque que cet indice ne s'applique que si le centre de  $\mathbf{A}$  est inférieur à celui de  $\mathbf{B}$ .

Nous utilisons la (légère) variante suivante, pour éliminer cette condition, et normaliser la mesure dans  $[0, 1]$  ([DARWICH et collab. \[2017\]](#)) :

$$\mathcal{A}'_{\ominus}(\mathbf{A}, \mathbf{B}) = \begin{cases} \mathcal{A}'_{\ominus}(\mathbf{A}, \mathbf{B}), & \text{si } c(\mathbf{A}) \leq c(\mathbf{B}) \text{ et } \mathcal{A}'_{\ominus}(\mathbf{A}, \mathbf{B}) \leq 1 \\ 0, & \text{si } c(\mathbf{A}) > c(\mathbf{B}) \\ 1, & \text{sinon.} \end{cases} \quad (3.9)$$

Finalement, la décision floue entre les hypothèses *Objet* et *Fond* du pixel  $p$ , est construite en comparant les intervalles  $\mathbf{F}_t(p)$  et  $\mathbf{B}_t(p)$  suivant cette variante de l'indice d'acceptabilité :

$$D_t(p) = \begin{cases} \frac{1}{2} \left[ 1 + \mathcal{A}'_{\ominus}(\mathbf{B}_t(p), \mathbf{F}_t(p)) \right], & \text{si } c(\mathbf{B}_t(p)) < c(\mathbf{F}_t(p)) \\ \frac{1}{2} \left[ 1 - \mathcal{A}'_{\ominus}(\mathbf{F}_t(p), \mathbf{B}_t(p)) \right], & \text{si } c(\mathbf{F}_t(p)) < c(\mathbf{B}_t(p)) \\ \frac{1}{2}, & \text{si } c(\mathbf{F}_t(p)) = c(\mathbf{B}_t(p)). \end{cases} \quad (3.10)$$

D'après cette formule, les valeurs supérieures (respectivement inférieures) à 0,5 induisent une décision *Objet* (r. *Fond*), alors que les valeurs proches de 0,5 indiquent juste un manque de préférence.

Pour compléter cette préférence, on lui associe un niveau de confiance  $E_t(p)$ . L'idée sous-jacente est de s'assurer qu'au moins un des 2 intervalles est suffisamment significatif. On s'en assure en comparant ces 2 intervalles à l'intervalle de rejet  $\mathbf{R}$  :

$$E_t(p) = \max \left\{ \mathcal{A}'_{\ominus}(\mathbf{R}, \mathbf{F}_t(p)), \mathcal{A}'_{\ominus}(\mathbf{R}, \mathbf{B}_t(p)) \right\}. \quad (3.11)$$

La figure 3.4 montre plusieurs paires d'intervalles *Objet* ( $F_t(p)$ ) et *Fond* ( $B_t(p)$ ), accompagnés de leurs mesures  $D_t(p)$  et  $E_t(p)$ . La surface grise représente la zone de rejet, définie par l'intervalle  $R$ .

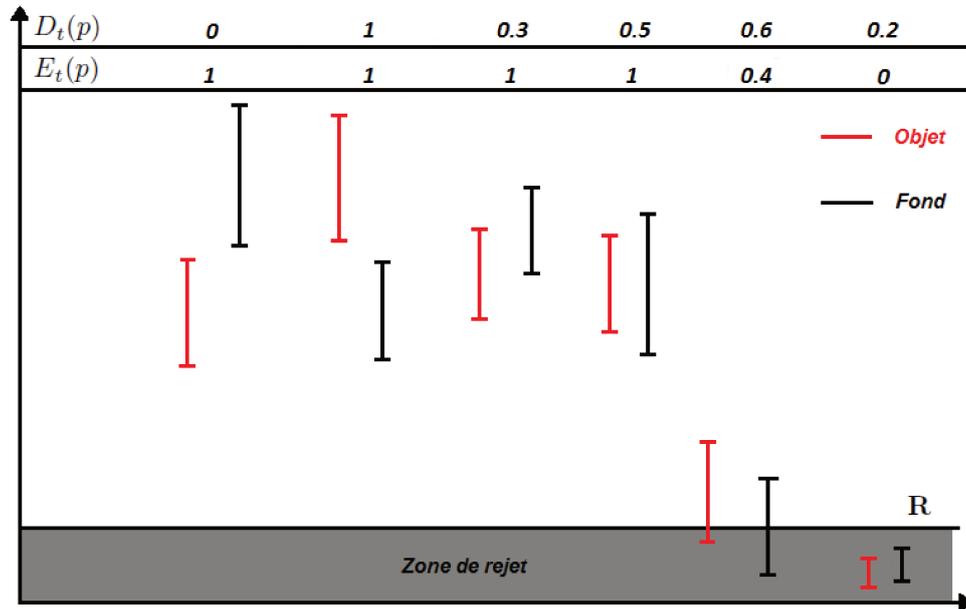


FIGURE 3.4 – Exemples de décisions floues par comparaison des intervalles agrégés.

### 3.3.3 Décision nette

À ce stade, il est possible de déterminer une décision nette à partir de la décision floue. Même s'il faut souligner que la gestion floue prend son sens dans un processus de fusion, avec des décisions complémentaires également floues (temporelles ou spatiales, comme nous le verrons par la suite).

La décision nette  $D_b(p)$  du pixel  $p$ , à valeurs binaires, peut être obtenue de la manière suivante :

$$D_b(p) = \begin{cases} 1, & \text{si } c(B_t(p)) < c(F_t(p)) \\ 0, & \text{sinon.} \end{cases} \quad (3.12)$$

## 3.4 Décisions complémentaires

En défuzzifiant la décision floue exposée ci-dessus, on peut obtenir une première estimation des pixels-objets. Mais dans une vidéo bruitée, ou avec un fond complexe (dynamique), certaines décisions risquent d'être ambiguës, par manque d'information.

On décide donc de conforter la décision en recherchant une cohérence spatiale ou temporelle.

### 3.4.1 Décision spatiale floue

On considère ici une hypothèse de continuité spatiale : des pixels *voisins* ont des chances d'être de la même classe. On définit un voisinage par une fenêtre  $W_s = 3 \times 3$  (par exemple).

Chaque pixel  $p$  dans l'image  $t$  est maintenant caractérisé par une décision floue  $D_t(p) \in [0,1]$  et un niveau de confiance  $E_t(p) \in [0,1]$ . La valeur de décision indique quelle classe *Objet* et *Fond* est préférée pour ce pixel. Et la valeur de confiance associée à cette décision établit sa contribution future à la décision spatiale.

Tout d'abord, la fenêtre  $W_s$  est centrée sur le pixel actuel  $p$  et remplie de ses similarités avec ses 8 voisins les plus proches. La similarité entre le pixel  $p$  et son voisin  $p'$  est définie à l'aide d'un noyau gaussien :

$$S(p,p') = e^{-\left(\frac{x_t^p - x_t^{p'}}{\sigma_S}\right)^2}, \quad (3.13)$$

avec  $\sigma_S$  un facteur d'échelle.

Les valeurs du masque sont ensuite pondérées par les degrés de confiance  $E_t(p')$  de tous les voisins  $p'$  de  $p$  (cf. figure 3.5).

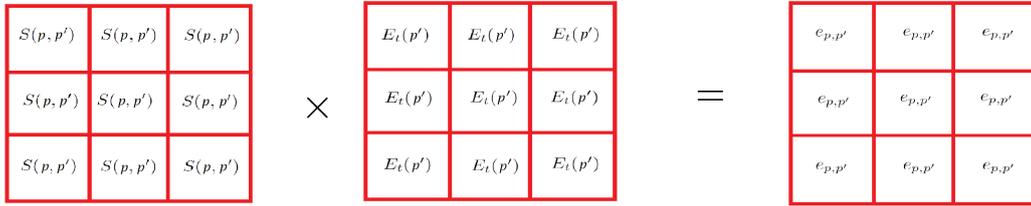


FIGURE 3.5 – Fusion spatiale : poids des décisions floues initiales.

Ensuite, ce masque est multiplié élément par élément par une matrice de même taille, composée des décisions floues intermédiaires du pixel voisin de  $p$  (cf. figure 3.6).

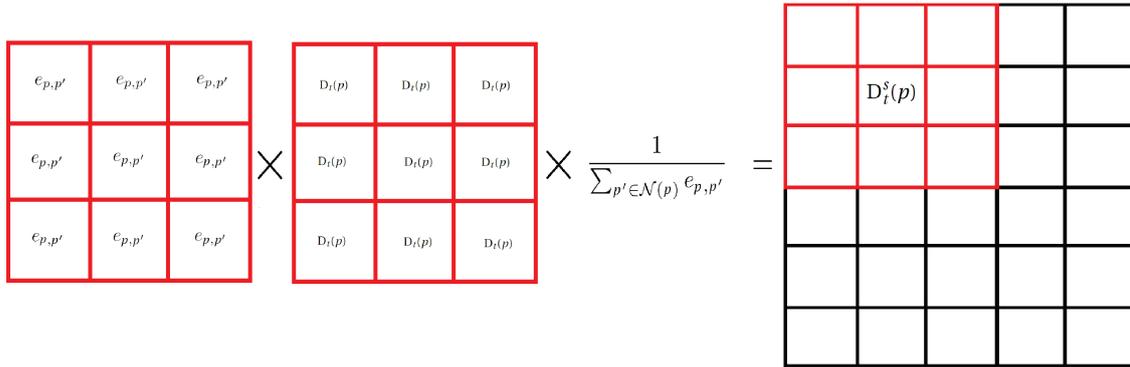


FIGURE 3.6 – Fusion spatiale : agrégation des décisions floues initiales.

La décision floue agrégée de  $p$  est donc définie comme une moyenne pondérée des termes de la matrice ainsi construite :

$$D_t^s(p) = \frac{1}{\sum_{p' \in \mathcal{N}(p)} e_{p, p'}} \sum_{p' \in \mathcal{N}(p)} e_{p, p'} \cdot D_t(p'), \quad (3.14)$$

avec :

- $\mathcal{N}(p)$  : les pixels voisins de  $p$ , définis par la fenêtre de voisinage  $W_s$  ;

- $e_{p,p'} = S(p,p').E_t(p')$  : le degré de contribution du pixel  $p'$  à la décision spatiale du pixel  $p$ .

Finalement, le niveau de confiance finale  $E_t^s(p)$  est définie comme la moyenne pondérée des niveaux de confiance du voisinage de  $p$  :

$$E_t^s(p) = \frac{1}{\sum_{p' \in \mathcal{N}(p)} S(p,p')} \sum_{p' \in \mathcal{N}(p)} S(p,p').E_t(p'). \quad (3.15)$$

### 3.4.2 Décision temporelle floue

Nous proposons d'intégrer un calcul de flot optique dans la méthode, pour exploiter la continuité temporelle des décisions. On a déjà présenté dans le premier chapitre des méthodes usuelles de calcul du flot optique. Pour résumer, elles estiment le champ de vitesses des pixels, en se basant sur l'hypothèse de conservation de la luminance (c'est-à-dire que la luminance d'un point physique de l'image - susceptible de se déplacer - ne varie pas au cours du temps).

Dans notre méthode, nous utilisons un algorithme d'estimation du flot optique efficace et robuste, conçu par Gunnar Farneback [FARNEBÄCK \[2003\]](#).

Nous proposons l'algorithme suivant, pour compléter l'information de classe du pixel  $p$  de l'image  $t$  :

- Calcul du flot optique pour chaque image  $t \neq 1$  (car la première image n'a pas d'image antécédente), de manière à associer à chaque pixel  $p$  son antécédent  $p'$  dans l'image précédente.
- Calcul de la confiance et de la décision du pixel par rapport à son antécédent :

$$\begin{aligned} E_t^{temp}(p) &= E_{t-1}^f(p'), \\ D_t^{temp}(p) &= D_{t-1}^f(p'), \end{aligned} \quad (3.16)$$

avec :

- $E_{f_{t-1}}(p')$ , le degré de confiance du pixel antécédent de  $p$ , de l'image  $t-1$ .
- $D_{f_{t-1}}(p')$ , la décision finale de ce pixel antécédent de  $p$ , de l'image  $t-1$ .

## 3.5 Fusion des décisions

À partir des deux décisions floues, spatiale et temporelle, il ne reste qu'à les fusionner pour construire la décision finale du pixel.

### 3.5.1 Décision finale floue

La décision finale, spatio-temporelle, est définie comme une moyenne pondérée :

$$\begin{aligned} D_f(p) &= \frac{D_t^s(p).E_t^s(p) + E_t^{temp}(p).D_t^{temp}(p)}{E_t^s(p) + E_t^{temp}(p)}, \\ E_f(p) &= \frac{1}{2} \left( E_t^s(p) + E_t^{temp}(p) \right), \end{aligned} \quad (3.17)$$

où  $D_f(p)$  et  $E_f(p)$  désignent respectivement la décision et le degré de confiance flou final du pixel  $p$ .

### 3.5.2 Défuzzification de la décision finale floue

La décision binaire est obtenue en arrondissant  $D_f(p)$  par l'entier le plus proche si le niveau de confiance final est suffisamment élevé, et en le fixant à 1 (*Objet*) sinon :

$$D_t^b(p) = \begin{cases} 0 \text{ (Fond)} & \text{si } D_f(p) \leq 0.5 \text{ et } E_f(p) > k_e \\ 1 \text{ (Objet)} & \text{si } D_f(p) > 0.5 \text{ et } E_f(p) > k_e \\ 1 & \text{si } E_f(p) \leq k_e. \end{cases} \quad (3.18)$$

Après cette étape de décision finale, une image binaire peut être calculée, pour marquer les emplacements des pixels *Objet*.

## 3.6 Autres étapes

Les étapes non détaillées jusqu'ici sont globalement similaires à celles de la méthode *TF2GMM* **BOUWMANS et EL BAF [2009]**. Nous soulignons ici les quelques différences.

### 3.6.1 Mise à jour du modèle *GMM*

Dans la méthode *TF2GMM* et celles voisines, la mise à jour du modèle *GMM* consiste à renforcer l'unique mode associé au pixel (celui le plus « vraisemblable »).

Dans notre méthode, nous procédons de la même manière, mais il faut d'abord identifier ce mode. En effet, la décision finale est issue de l'agrégation de plusieurs décisions complémentaires, conçues à partir de différents modèles *GMM* (suivant le pixel ou l'image), et aussi de plusieurs modes (les intervalles *Fond* et *Objet* sont obtenus par agrégation des intervalles provenant de tous les modes).

En pratique, nous recherchons donc, pour le pixel considéré  $p$ , dans l'image  $t$ , le mode  $k^*$  le plus vraisemblable de la même classe (*Objet* ou *Fond*) :

$$k^* = \arg \max_{k \in \{1, \dots, K_B\}} f_k(x_t^p), \quad \text{si le pixel est finalement classé } \textit{Fond}, \quad (3.19)$$

$$= \arg \max_{k \in \{K_B+1, \dots, K\}} f_k(x_t^p), \quad \text{sinon.} \quad (3.20)$$

On considère que la sélection a échoué si la vraisemblance du mode optimal est inférieure au seuil de rejet  $R$ . En ce cas, un nouveau mode gaussien remplace celui le moins prioritaire (de poids  $\frac{w_k}{\sigma_k}$  le plus faible), centré sur le niveau du pixel  $x_t^p$ , avec un poids initialisé à  $\beta$ .

Une fois la mise à jour effectuée, l'algorithme reprend en début de boucle, par l'étape de fusion des modes proches (cf. algorithme 3.1).

### 3.6.2 Fusion des modes proches

On a proposé cette étape pour remédier aux modes de centres trop proches (gaussiennes qui se superposent).

L'étape se divise en deux parties :

- Calcul de la similarité  $S(a,b)$  entre chaque paire de modes  $(a,b)$  : il s'agit simplement d'évaluer la vraisemblance du centre  $\mu_b$  du mode  $b$ , par la fonction de vraisemblance du mode  $a$ , et réciproquement (cf. équation 1.10 pour la vraisemblance  $RGB$ ); puis de retenir le maximum de ces 2 valeurs :

$$S(a,b) = \max\left(f_g\left(\mu_{a,t}^p; \mu_{b,t}^p, \Sigma_{b,t}^p\right), f_g\left(\mu_{b,t}^p; \mu_{a,t}^p, \Sigma_{a,t}^p\right)\right). \quad (3.21)$$

- Chaque paire de modes dont la similarité est suffisamment forte ( $S(a,b) > k_s$ ) est fusionnée en un unique nouveau mode, suivant la formule proposée par SHIMADA et collab. [2006] :

$$\begin{aligned} w &= w_a + w_b; \\ \mu &= \frac{w_a * \mu_a + w_b * \mu_b}{w_a + w_b}; \\ \sigma^2 &= \frac{w_a * \sigma_a^2 + w_b * \sigma_b^2}{w_a + w_b}. \end{aligned} \quad (3.22)$$

## 3.7 Conclusion du chapitre

Dans ce chapitre nous avons détaillé notre algorithme, en montrant comment exploiter un mélange de gaussiennes floues, pour construire une décision d'abord floue, puis finalement nette.

Une caractéristique essentielle de la méthode est de maintenir la gestion de l'incertitude - introduite par les IVFS au niveau des modes gaussiens - tout au long de la procédure conduisant à la décision finale. Cette décision est par conséquent établie en accordant à chaque « canal » d'information (pixel, spatial, ou temporel) une influence fonction de la confiance qu'on peut lui prêter. Cette confiance est directement construite à partir de la part d'incertitude - imprécision - mesurée.

Par ailleurs, contrairement aux méthodes type  $GMM$  rencontrées jusque-là, les mesures d'incertitudes sont utilisées pour attribuer un degré de contribution à chaque mode gaussien. Alors que dans les méthodes citées, les décisions sont construites à partir d'un unique mode, celui le plus vraisemblable. En préférant faire participer tous les modes au processus de décision, par une pondération, on obtient par conséquent une décision moins arbitraire.

Ces 2 aspects permettent la construction d'une décision « finale » jugée plus pertinente que celle des méthodes  $GMM$  classiques, des méthodes basées sur un  $GMM$  ou sur un formalisme flou.

Dans le chapitre suivant nous allons évaluer l'impact de chacune des étapes, puis comparer les performances de la méthode avec un ensemble de méthodes issues de la littérature.

## 3.8 Références

BOUWMANS, T. et F. EL BAF. 2009, «Modeling of dynamic backgrounds by type-2 fuzzy gaussian mixture models», *MASAUM Journal of Basic and Applied Sciences*, vol. 1, n° 2, p. 265–276. 63, 71

- DARWICH, A., P.-A. HÉBERT, Y. MOHANNA et A. BIGAND. 2017, «Background subtraction under uncertainty using a type-2 fuzzy set gaussian mixture model», dans *The Fourth International Conference on Computer Science, Computer Engineering, and Education Technologies (CSCEET2017)*, IEEE, p. 1–6. [67](#)
- EL BAE, F., T. BOUWMANS et B. VACHON. 2008, «Type-2 fuzzy mixture of gaussian model : application to background modeling», dans *International Symposium on Visual Computing*, Springer, p. 772–781. [63](#)
- FARNEBÄCK, G. 2003, «Two-frame motion estimation based on polynomial expansion», *Image analysis*, p. 363–370. [70](#)
- SENGUPTA, A. et T. K. PAL. 2000, «On comparing interval numbers», *European Journal of Operational Research*, vol. 127, n° 1, p. 28–43. [67](#)
- SHIMADA, A., D. ARITA et R.-I. TANIGUCHI. 2006, «Dynamic control of adaptive mixture-of-gaussians background model», dans *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*, IEEE, p. 5–5. [72](#)
- STAUFFER, C. et W. E. L. GRIMSON. 1999, «Adaptive background mixture models for real-time tracking», dans *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, IEEE. [63](#), [64](#), [65](#)



# Chapitre 4

## Validation de la méthode

### Sommaire

---

<b>4.1 Introduction</b> . . . . .	<b>76</b>
<b>4.2 Indicateurs de performance</b> . . . . .	<b>76</b>
4.2.1 Classement des décisions . . . . .	76
4.2.2 Définition des indicateurs de performance . . . . .	77
<b>4.3 Validation des étapes de la méthode</b> . . . . .	<b>77</b>
4.3.1 Protocole d'évaluation . . . . .	77
4.3.2 Analyse des résultats . . . . .	79
<b>4.4 Comparaison à des méthodes basées sur le <i>GMM</i> et le flou</b> . . . . .	<b>82</b>
4.4.1 Base de vidéos de référence . . . . .	82
4.4.2 Protocole d'évaluation . . . . .	83
<b>4.5 Comparaison à des méthodes basées sur l'<i>ACP</i></b> . . . . .	<b>86</b>
<b>4.6 Comparaison des temps de calcul des différentes méthodes</b> . . . . .	<b>87</b>
<b>4.7 Conclusion du chapitre</b> . . . . .	<b>89</b>
<b>4.8 Références</b> . . . . .	<b>90</b>

---

## 4.1 Introduction

Après l'état de l'art des deux premiers chapitres, décrivant des méthodes usuelles d'extraction de l'arrière plan, nous avons présenté dans le chapitre précédent, une nouvelle méthode basée sur les ensembles flous de type-2 (valeurs d'appartenance de type intervalle), particulièrement robuste aux changements dynamiques d'arrière-plan.

Ce chapitre présente les résultats d'un ensemble d'expérimentations. Une première partie de ces tests a permis de montrer l'intérêt de la méthode pour la soustraction d'arrière-plans dynamiques, avec en particulier l'évaluation de l'impact des différentes étapes du processus de décision.

La seconde partie des expérimentations a permis de la comparer à celles de la littérature, précédemment décrites.

Tous les calculs d'évaluation quantitative de notre méthode ont été lancés sur le serveur **CALCULCO** de l'Université du Littoral. Dans un premier temps, nous avons développé un prototype de notre méthode sous MATLAB™. Les temps de calcul importants nous ont ensuite orienté vers une programmation en C++, avec la bibliothèque *OpenCV*.

Les évaluations des méthodes comparées ont été extraites du site **CDNET**, lorsqu'elles étaient disponibles.

Les résultats restants ont été obtenus soit avec la librairie *LRSLibrary* **SOBRAL et collab. [2015]** (méthodes d'ACP robustes), soit avec la librairie *BGSlibrary* **SOBRAL [2013]** (méthodes de soustraction de fond), avec les paramètres par défaut, sur un ordinateur équipé d'un *CPU Intel(R) Core(TM) i7-4500UCPU@1.80Ghz2.40Ghz*, avec 16Go de RAM.

## 4.2 Indicateurs de performance

### 4.2.1 Classement des décisions

Chaque pixel fait l'objet d'une décision, que l'on peut classer dans l'une de ces 4 catégories :

	Décision	
Vérité	<i>Objet</i>	<i>Fond</i>
<i>Objet</i>	TP	FN
<i>Fond</i>	FP	TN

TABLEAU 4.1 – Classement des décisions.

Sur l'ensemble des images d'une vidéo, les effectifs de ces 4 classes donnent une indication des performances de la procédure de reconnaissance *Objet/Fond* :

- TP (True Positive) : nombre de *vrais positifs* (ie. décisions *Objet* correctes).
- FP (False Positive) : nombre de *faux positif* (ie. fausses décisions *Objet*).
- TN (True Negative) : nombre de *vrais négatif* (ie. décisions *Fond* correctes).
- FN (False Negative) : nombre de *faux négatifs* (ie. fausses décisions *Fond*).

## 4.2.2 Définition des indicateurs de performance

Des indicateurs de performances plus synthétiques sont définis comme des combinaisons de ces effectifs (cf. Tableau 4.2).

Paramètre	Formule
Rappel (ou taux de détection)	$R_e = \frac{TP}{TP+FN}$
Précision	$P_r = \frac{TP}{TP+FP}$
Spécificité	$S_p = \frac{TN}{TN+FP}$
Ratio de classifications erronées	$PBC = \frac{FN+FP}{TP+FN+FP+TN}$
F-mesure	$F\text{-mesure} = \frac{2 \cdot P_r \cdot R_e}{P_r + R_e}$

TABLEAU 4.2 – Indicateurs de performance.

Ces paramètres usuels ont les significations suivantes :

- $R_e$  : proportion de pixels de l'arrière-plan correctement identifiés.
- $P_r$  : proportion de pixels du premier plan correctement identifiés.
- $F_m$  : moyenne géométrique du Rappel et de la Précision.
- PBC : proportion des erreurs de décision.
- $S_p$  : proportion des pixels de l'arrière-plan correctement identifiés.

## 4.3 Validation des étapes de la méthode

### 4.3.1 Protocole d'évaluation

#### Optimisation des paramètres

Pour valider les différentes étapes de la méthode, nous avons travaillé sur une seule vidéo de référence : la vidéo *Canoé*. Issue de la catégorie de vidéos *Dynamic background* (depuis le site [CDNET](#)), elle montre le passage d'un bateau sur une rivière, aux reflets changeants. C'est une petite vidéo (dimensions 320 x 240), avec 30 images par seconde, et 1189 images au total.

Nous l'avons sélectionnée, car c'est une vidéo type de sa catégorie, qui n'offre normalement pas de difficulté particulières aux algorithmes de soustraction de fond.

Les paramètres de notre méthode ont été optimisés, pour l'adapter au mieux à cette vidéo. La liste des paramètres est rappelée dans le tableau 4.3.

Tous ces paramètres doivent dans l'idéal être judicieusement fixés, en fonction de la vidéo. Dans de nombreuses applications, il est en effet possible de procéder à une optimisation de leurs valeurs. On obtient alors de meilleurs scores qu'avec des paramètres « par défaut ».

Paramètre	Valeur
$K$	Nombre de modes (gaussiennes)
$\alpha$	Taux d'apprentissage (learning rate)
$\beta$	Poids des nouveaux modes
$T_b$	Seuil de fond
$k_m$	Quantité de FOU
$k_r$	Borne supérieure de l'intervalle de rejet
$k_e$	Seuil du degré de confiance
$k_s$	Seuil de la vraisemblance
$\sigma_n$	L'écart-type des nouveaux modes
$W_s$	Taille de la fenêtre de voisinage

TABLEAU 4.3 – Paramètres de la méthode

### Procédure

Nous avons optimisé ces paramètres de manière à assurer la meilleure reconnaissance des objets, que nous avons évaluée par l'indicateur - synthétique - de F-mesure.

Voici ci-dessous la procédure suivie.

- Génération d'un ensemble de jeux de paramètres (2443 jeux, obtenus par combinaison de valeurs de paramètres choisies de manière heuristique). Chaque jeu de paramètres attribue une combinaison de valeurs unique à l'ensemble des paramètres.
- Sélection du meilleur jeu de paramètres, selon la variante complète (spatio-temporelle) : ce jeu est obtenu en maximisant la F-mesure (cf. tableau 4.4).

Paramètre	Valeur
$K$	3
$\alpha$	0,01
$\beta$	0,001
$T_b$	0,85
$k_m$	2
$k_r$	0,07
$k_e$	0,25
$k_s$	0,75
$\sigma_n$	36
$W_s$	5

TABLEAU 4.4 – Jeu de paramètres optimisé pour la vidéo canoé

- Lancement des calculs avec le jeu de paramètre optimisé, avec plusieurs variantes de notre méthode *IV-FGMM* (obtenues en activant/désactivant certaines étapes) :

- Décision binaire : la comparaison des intervalles est réduite à la comparaison nette de leurs centres.  
Les variantes concernées sont suffixées par la lettre « C » pour *Crisp* (la traduction de « net » qui s’oppose à *fuzzy* dans la théorie des sous-ensembles flous).
- Décision floue : la comparaison des intervalles est opérée de la manière décrite au chapitre 3, dans le but d’obtenir une mesure de décision floue entre les classes *Fond* et *Objet*, ainsi qu’un degré de confiance dans cette préférence.
- Décision spatiale : fusion des décisions floues des pixels voisins.  
Les variantes concernées sont suffixées par la lettre « S » pour *Spatial*.
- Décision temporelle : fusion de la décision floue de l’image  $t$ , avec une décision temporelle floue construite à partir de la décision finale de l’image précédente  $t - 1$  (grâce à une estimation du mouvement des pixels, issue d’un calcul de flot optique).  
Les variantes concernées sont suffixées par la lettre « T » pour *Temporal*.
- Décision spatio-temporelle : fusion des décisions spatiales et temporelles décrites précédemment.  
Les 2 variantes concernées (floue et nette) sont donc caractérisées par « ST ».

### 4.3.2 Analyse des résultats

Le tableau 4.5 présente les performances des différentes variantes de notre méthode sur la vidéo *Canoé*, obtenues avec le jeu de paramètres optimisé.

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>Binaire</b>	23,770	34,294	2,431	939,504	0,965	3,678	0,907	0,410	<b>0,564</b>
Binaire/temporelle	23,469	18,861	2,741	954,929	0,981	2,163	0,895	0,554	0,685
Binaire/spatiale	20,145	0,171	3,565	976,118	1	0,375	0,849	0,992	0,915
<b>Binaire/spatio-temporelle</b>	20,038	0,146	3,794	976,023	1	0,393	0,841	0,993	<b>0,911</b>
<b>Flou</b>	22,651	30,915	2,646	943,787	0,968	3,348	0,895	0,423	<b>0,574</b>
Flou/temporelle	23,084	20,461	2,551	953,903	0,979	2,296	0,900	0,530	0,667
Flou/spatial	22,172	0,383	2,416	975,029	1	0,280	0,902	0,983	0,941
<b>Flou/spatio-temporelle</b>	2,2233	0,260	2,419	975,087	1	0,268	0,902	0,988	<b>0,943</b>

TABLEAU 4.5 – Performances de chaque variante, vidéo Canoé.

#### Intérêt du spatial et du flou

On constate que c’est surtout l’intégration de l’information spatiale qui améliore le processus de décision (F-mesure de 0,911 pour la variante binaire/spatiale contre 0,564 pour la variante binaire).

L’ajout de la gestion floue permet d’accroître les performances (0,941 pour la variante floue/temporelle).

Pour se rendre compte - plus qualitativement - de l’importance des différentes étapes de la méthode, nous montrons maintenant quelques résultats de classification sur une image de la séquence : l’image 965, où l’on peut observer le passage du canoé (cf. les 2 images supérieures de la figure 4.1, montrant l’image 4.1a et sa vérité-terrain attendue 4.1f).

Les images de la figure 4.1 permettent de visualiser les résultats de la détection d'objets, selon les différentes variantes étudiées. On peut observer l'intérêt de l'étape de fusion spatiale/floue : elle élimine beaucoup d'erreurs de classification.

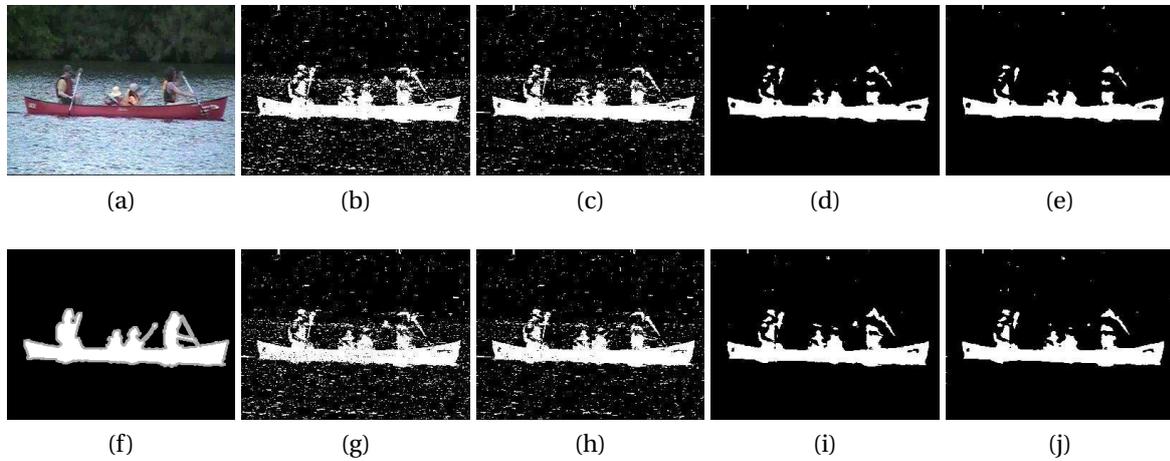


FIGURE 4.1 – Image initiale (a), Binaire (0.564) (b), Binaire/temporelle (0.685) (c), Binaire/spatiale(0.915)(d), Binaire/spatio-temporelle(0.911)(e), Vérité terrain(f), Floue(0.574)(g), Floue/temporelle(0.667)(h), Floue/spatiale(0.9)(i), Floue/spatio-temporelle(0.943), (j).

L'apport du temporel est moins évident (0,943 pour la variante floue/spatio-temporelle). Il faudrait sans doute complexifier un peu l'exploitation du flot optique, pour en retirer un gain notable.

### Analyse des indicateurs de performance

La figure 4.2 fournit les mesures de spécificité de chaque variante de la méthode. On constate que c'est la fusion spatiale qui permet d'atteindre la spécificité maximale (1).

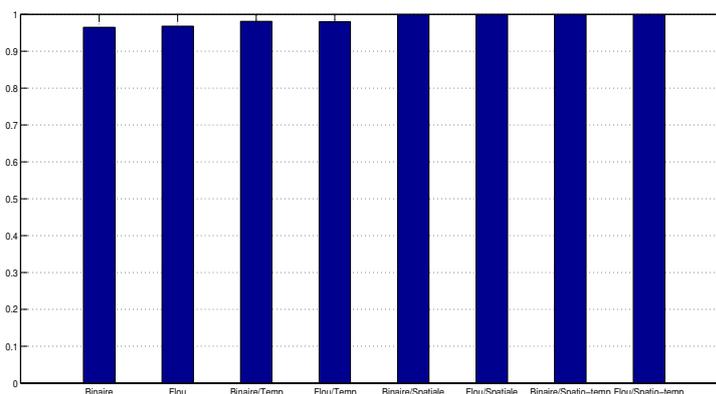


FIGURE 4.2 – Spécificité selon les variantes, vidéo Canoé.

La figure 4.3 présente la proportion d'erreurs de classification. Le variante la moins performante est celle nette (binaire) sans fusion. C'est à l'inverse les variantes spatiales/floues qui atteignent les meilleurs résultats (0.268% pour la version floue spatiale et temporelle).

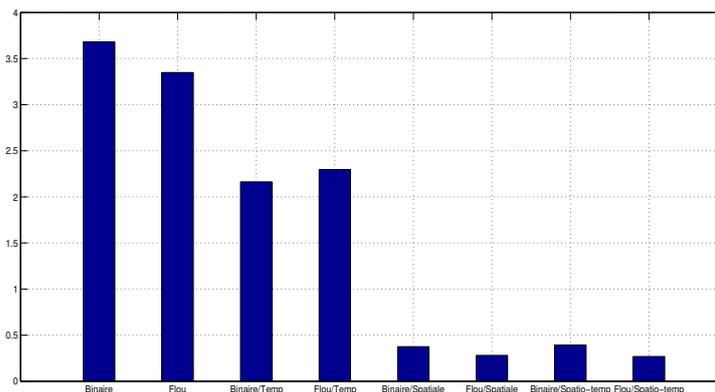


FIGURE 4.3 – PCB selon les variantes, vidéo Canoé.

Les mesures de rappel (figure 4.4) sont bonnes, et proches les unes des autres, exceptées pour les variantes binaires spatiales (temporelles ou non).

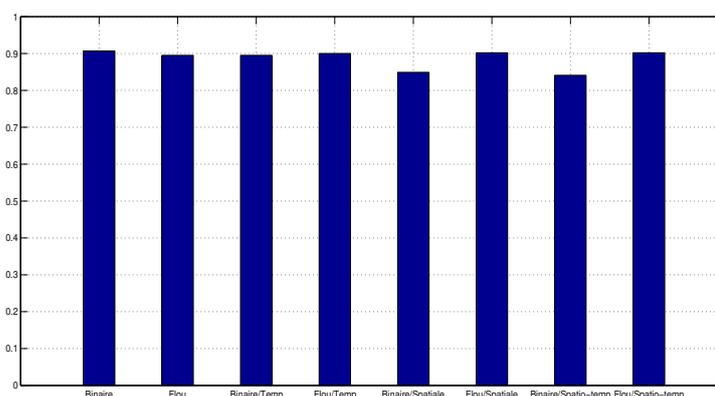


FIGURE 4.4 – Rappel selon les variantes, vidéo Canoé.

La figure 4.5 démontre clairement l'utilité de l'étape de fusion spatiale : toutes les variantes sans cette fusion obtiennent de nettement moins bons scores.

Enfin, le tableau des F-mesures (cf. figure 4.6) - mesure synthétique combinant le rappel et la précision - montre l'intérêt de la fusion spatiale, alliée à une gestion des incertitudes : ce sont les variantes flou/spatial et flou/spatio-temporel qui atteignent les plus haut scores. Même si la contribution du flou à la F-mesure n'est pas très importante (+2%), sur la quantité de pixels testés, cela reste tout de même intéressant.

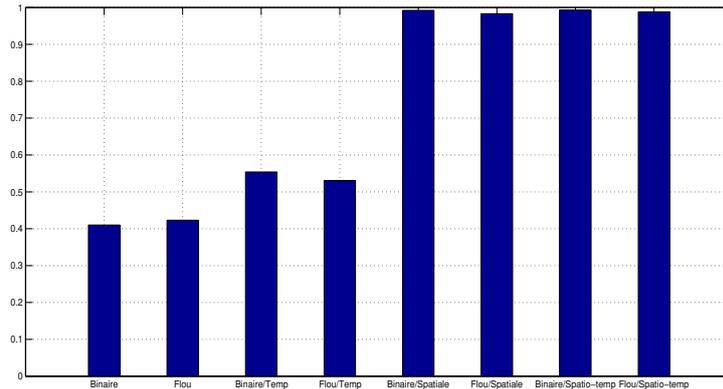


FIGURE 4.5 – Précision selon les variantes, vidéo Canoé.

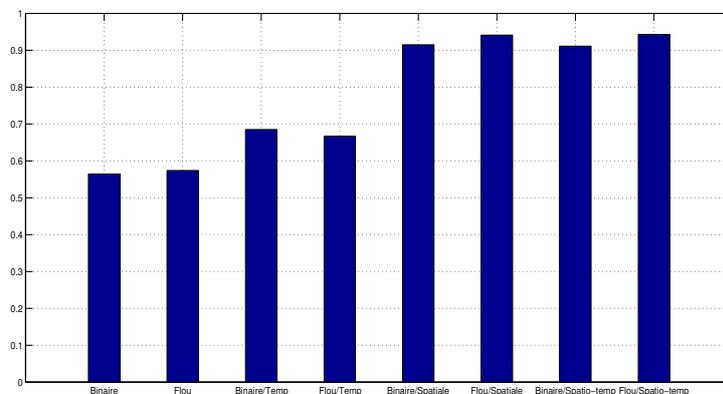


FIGURE 4.6 – F-mesures selon les variantes, vidéo Canoé.

## 4.4 Comparaison à des méthodes basées sur le *GMM* et le flou

### 4.4.1 Base de vidéos de référence

En 2012 puis 2014, les auteurs de la communication [WANG et collab. \[2014\]](#) ont soumis une « base de vidéos de référence » (sur le site [CDNET](#)), un ensemble de vidéos destiné à l'évaluation des algorithmes de détection d'objets, dans des situations difficiles (mais réalistes). Cette base constitue une référence académique rigoureuse et complète largement utilisée pour la comparaison des méthodes.

Le site propose deux bases de vidéos : *DATASET 2012* et *DATASET 2014* :

- La base de 2012 est composée 31 séquences vidéo, rassemblant un total de près de 90000 images. Ces vidéos sont regroupées en 6 catégories : fond dynamique, gigue de caméra, Objet de premier plan endormi, ombres et signatures thermiques cf. section 1.2.
- En 2014 la base est complétée avec 5 nouvelles catégories : météo difficile, faible taux de rafraîchissement, acquisition nocturne (changements de luminosité), Réglages automatiques de la caméra (PTZ), et turbulence de l'air cf. section 1.2. Au total, plus de 70000 images ont été capturées, puis

segmentées et annotées manuellement par une équipe de 13 chercheurs de 7 universités **WANG et collab. [2014]**.

La base de données complète (2014) fournit un ensemble de vidéos réalistes et difficiles à traiter, capturées par caméra. Elles sont représentatives des scènes d'intérieur et d'extérieur capturées aujourd'hui dans des scénarios de surveillance, d'environnement intelligent et de base de données vidéo.

Pour évaluer notre méthode, nous nous sommes exclusivement intéressés à la catégorie des vidéos à fonds dynamiques : c'est en effet la plus pertinente pour mesurer la capacité de notre modèle flou à s'adapter à des arrière-plans mouvants ou changeants.

#### 4.4.2 Protocole d'évaluation

##### Optimisation de paramètres dans la catégorie de vidéos « *fond dynamique* »

Pour évaluer notre méthode, nous avons utilisé le jeu de paramètres maximisant la F-mesure moyenne ( $F_m$  dans les tableaux) de la version la plus complète de notre méthode (soit celle spatio-temporelle floue), sur l'ensemble des vidéos de la catégorie *fond dynamique*.

En cela, nous n'avons pas exactement appliqué le protocole officiel du site *ChangeDetection* : il aurait fallu optimiser la F-mesure moyenne sur les vidéos de *toutes* les catégories. Mais, vu l'importante quantité de combinaisons de paramètres à tester, nous avons dû nous restreindre à la catégorie qui nous intéressait le plus (*fond dynamique*).

Le jeu de paramètres optimisé est présenté dans le tableau 4.6.

Paramètre	Valeur
$K$	3
$\alpha$	0,01
$\beta$	0,04
$T_b$	0,85
$k_m$	2,5
$k_r$	0,05
$k_e$	0,25
$k_s$	0,75
$\sigma_n$	36
$W_s$	5

TABLEAU 4.6 – Jeu de paramètres optimisé sur la catégorie de vidéos *fond dynamique*.

#### Méthodes utilisées dans la comparaison

Nous avons ensuite comparé notre méthode à des méthodes à base de *GMM* ou de modélisation floue.

On a également comparé l'algorithme à des méthodes basées sur une réduction de dimension (*ACP* classique ou robuste). Les *ACP* robustes sont en effet des techniques récentes, particulièrement adaptées au problème d'extraction du fond en vidéo.

L'ensemble des méthodes sélectionnées sont présentées dans le tableau 4.7 ci-dessous.

Méthode	Référence
GMM-Zivkovic (GMM-Z)/CDNET	ZIVKOVIC [2004]
KDE-El-Gammal (KDE-G)/CDNET	ELGAMMAL et collab. [2000]
GMM-Stauffer-Grimson (GMM-SG)/CDNET	STAUFFER et GRIMSON [1999]
RMoG (Region-based MoG)/CDNET	VARADARAJAN et collab. [2013]
BMOG/CDNET	MARTINS et collab. [2017]
T2-FGMM-UM/BGS	EL BAF et collab. [2008]
T2-FGMM-UV/BGS	EL BAF et collab. [2008]
T2-FGMM-UMRF/BGS	ZHAO et collab. [2012]
T2-FGMM-VMRF/BGS	ZHAO et collab. [2012]
Fuzzy Sugeno Integral (FSI)/BGS	ZHANG et XU [2006]
Fuzzy Choquet Integral (FCI)/BGS	BAF et collab. [2008]
Fuzzy Gaussian (FG)/BGS	SIGARI et collab. [2008]
ACP Classique	DARWICH et collab. [2015]
RACP/LRS	DE LA TORRE et BLACK [2001]
FPCP/LRS	RODRÍGUEZ et WOHLBERG [2013]
IALM/LRS	LIN et collab. [2010]
EALM/LRS	LIN et collab. [2010]
ALM/LRS	TANG et NEHORAI [2011]
LSADM/LRS	GOLDFARB et collab. [2013]
MoG-RPCA/LRS	ZHAO et collab. [2014]

TABEAU 4.7 – Méthodes utilisées dans l'étude comparative.

## Résultats de la comparaison des méthodes

Les résultats des méthodes *GMM-GS*, *GMM-Z*, *KDE-G*, *RMoG*, et *BMOG* sont directement extraits du site [CDNET 2014](#).

Les résultats des méthodes de type *ACP* robuste ont été calculés, grâce à la librairie *LRSLibrary* [SOBRAL et collab. \[2015\]](#).

Quant aux autres méthodes, nous avons utilisé la librairie *BGSLibrary* [SOBRAL \[2013\]](#), avec les paramètres par défaut de chaque méthode.

Nous avons fait une exception avec la méthode *T2-FGMM* : les scores étant étonnamment faibles, nous avons essayé de les optimiser. Nous nous attendions en effet à des performances plus proches des nôtres, en raison des similitudes entre les 2 méthodes.

Les paramètres ont été obtenus après maximisation de la F-mesure sur toutes les vidéos de la catégorie (sur un ensemble de jeux de paramètres définis de manière heuristique). Les combinaisons obtenues pour

Méthode	Paramètres			
	$k_m$	$k_v$	$\alpha$	$k_p$
T2-FGMM-UM	1,9	-	0,001	6
T2-FGMM-UV	-	0,95	0,001	40
T2-FGMM-UMRF	1,9	-	0,001	5
T2-FGMM-VMRF	-	0,95	0,001	40

TABLEAU 4.8 – Paramètres optimisés pour les méthodes *T2-FGMM*.

chacune des 3 variantes de *T2-FGMM* basées sur le mélange de gaussiennes floues de type-2, sont données dans le tableau 4.8.

Le tableau 4.9 rassemble l'ensemble des résultats utiles à la comparaison des méthodes.

Méthode	Fall	Canoé	Overpass	Fontaine1	Fontaine2	Boats	$F_m$
<b>IV-FGMM-ST</b>	0,637	0,920	0,891	0,075	0,939	0,570	<b>0,672</b>
GMM-Z	0,423	0,885	0,867	0,081	0,791	0,747	0,632
KDE-G	0,308	0,882	0,824	0,105	0,823	0,632	0,596
GMM SG	0,435	0,881	0,871	0,076	0,803	0,728	0,633
<b>RMoG</b>	0,673	0,935	0,901	0,203	0,865	0,832	<b>0,735</b>
<b>BMOG</b>	0,691	0,950	0,962	0,381	0,932	0,838	<b>0,792</b>
T2-FGMM-UM	0,065	0,129	0,419	0,031	0,485	0,053	0,197
T2-FGMM-UV	0,015	0,208	0,384	0,056	0,529	0,011	0,200
T2-FGMM-UMRF	0,065	0,194	0,390	0,021	0,479	0,058	0,201
T2-FGMM-VMRF	0,015	0,203	0,374	0,051	0,502	0,009	0,192
FSI	0,225	0,820	0,474	0,068	0,508	0,379	0,201
FCI	0,215	0,836	0,403	0,051	0,533	0,429	0,411
FG	0,072	0,131	0,092	0,005	0,010	0,018	0,056

TABLEAU 4.9 – F-mesures de l'ensemble des vidéos.

### Analyse des résultats

Dans ce tableau 4.9, ainsi que dans les résultats plus détaillés en annexe A.1, on observe tout d'abord que notre méthode donne des résultats comparables, sinon meilleures, aux autres méthodes, y compris celles basées sur un modèle *GMM* - flou ou non -, à l'exception des méthodes *BMOG* et *RMoG*.

Par rapport à notre méthode, *RMoG* exploite davantage le voisinage des pixels : déjà dans l'étape d'initialisation, ensuite dans celle de mise à jour. Quant à la méthode *BMOG*, elle utilise un espace couleur plus approprié que le classique *RGB - CIE L\*a\*b* - ainsi qu'un mécanisme de taux d'apprentissage dynamique.

En dehors de ces méthodes, qui exploitent des informations de nature différentes, notre méthode obtient donc de bons résultats.

En particulier, elle atteint ses meilleurs scores avec la vidéo *fontaine2*, avec une F-mesure de 0,939, supérieure à celle des autres méthodes. C'est pourtant une vidéo relativement complexe, à cause du mouvement des particules d'eau provenant de la grande fontaine.

On constate que la vidéo *fountain1* A.5 pose problème à l'ensemble des méthodes. Sa complexité hors-norme est due à l'allure changeante de ses 4 fontaines, à la très faible taille de l'objet, ainsi qu'à sa couleur, très proche de celle du fond.

Mais il faut quand même relativiser ces performances, compte tenu de l'optimisation du jeu de paramètres de notre méthode, spécifique à cette catégorie *fond dynamique*.

## 4.5 Comparaison à des méthodes basées sur l'ACP

Ces méthodes basées sur une réduction de dimensions sont naturellement gourmandes en temps de calcul, comme en quantité de mémoire. Pour rendre les calculs possibles, nous avons dû réduire la taille des images des vidéos (vidéos tests + vérité-terrain), en utilisant un filtre médian.

Pour notre méthode, nous avons procédé à une nouvelle optimisation du jeu de paramètres sur ces vidéos réduites (cf. tableau 4.10).

Paramètre	Valeur
$\alpha$	0,01
$\beta$	0,01
$T_b$	0,85
$k_m$	3
$k_r$	0,01
$k_e$	0,25
$k_s$	0,75
$\sigma_n$	36
$W_s$	5

TABLEAU 4.10 – Paramètres optimisés sur les vidéos réduites.

### Analyse de résultats

Le tableau suivant (4.11) présente les performances des différents algorithmes basés sur l'ACP, ainsi que dans les résultats plus détaillés en annexe A.2 pour l'ensemble des vidéos testées.

On constate tout d'abord que les résultats diffèrent beaucoup selon la vidéo traitée : la vidéo *overpass* réussit à la plupart des méthodes, là où la vidéo *fountain1* les met toutes en échec. Sur les 4 autres vidéos, les résultats sont cependant plus homogènes, et assez satisfaisants.

Notre méthode échoue seulement sur la vidéo *boats*, avec un score  $F_m = 0,14$ , bien inférieur à celui des autres méthodes, qui atteignent des valeurs entre 0,65 et 0,72. Le jeu de paramètre optimisé « globalement » n'est pas du tout adapté à cette vidéo (en effet, avec un jeu de paramètres optimisé, le score s'élève alors à 0,695).

Elle se rattrape en revanche avec la vidéo *fall*, avec un excellent score  $F_m = 0,928$ , là où les autres méthodes stagnent plutôt entre 0,5 et 0,7.

Méthode	Fall	Canoé	Overpass	Fontaine1	Fontaine2	Boats	$F_m$
<b>IV-FGMM-ST</b>	0,928	0,834	0,877	0,197	0,909	0,143	<b>0,648</b>
ACP-classique	0,503	0,384	0,663	0,04	0,656	0,658	0,484
RACP	0,481	0,469	0,786	0,036	0,685	0,658	0,519
LSADM	0,525	0,446	0,821	0,052	0,742	0,7	0,548
ALM	0,525	0,892	0,861	0,029	0,591	0,7	0,6
EALM	0,525	0,471	0,809	0,047	0,593	0,7	0,524
IADM	0,683	0,471	0,808	0,047	0,667	0,719	0,566
FPCP	0,525	0,892	0,861	0,029	0,591	0,7	0,6
MOG-ACP	0,609	0,4	0,855	0,036	0,676	0,708	0,547

TABLEAU 4.11 – F-mesures des méthodes de type ACP

Globalement, sur l'ensemble des vidéos, les résultats sont meilleurs que les autres méthodes : on obtient la plus forte F-mesure moyenne ( $F_m = 0,648$ ).

Parmi les méthodes d'ACP, deux variantes robustes se distinguent néanmoins : *FPCP* et *ALM* (F-mesures moyennes à 0,6). La première est intéressante car elle est plus économe en mémoire et calcul que les autres variantes.

## 4.6 Comparaison des temps de calcul des différentes méthodes

Étant donnée la complexité algorithmique de certaines méthodes, nous avons effectué nos comparaisons sur une vidéo « allégée ». Nous l'avons constituée en extrayant une séquence d'images consécutives de la vidéo *fontaine1* (entre 50 et 5016, avec 30 images par seconde), puis en les réduisant aux dimensions  $160 \times 128$ .

Nous avons ensuite testé toutes les méthodes (à l'exception de *RMoG* et *BMOG*, que nous n'avons pas eu le temps de développer) sur cette séquence vidéo, avec un ordinateur équipé d'un *CPU Intel(R) Core(TM) i7 – 4500UCPU@1.80Ghz2.40Ghz*, avec 16GB de RAM.

Pour notre méthode, nous avons comparé plusieurs versions :

- Avec parallélisation sur 4 cœurs (leur nom est suffixé par **-P**), et sans parallélisation (sans suffixe).
- La méthode avec décision flou (*IV-FGMM*, *IV-FGMM-P*), avec décision temporelle seule (*IV-FGMM-T*, *IV-FGMM-T-P*), avec décision spatiale seule (*IV-FGMM-S*, *IV-FGMM-S-P*), et la version complète spatio-temporelle (*IV-FGMM-ST*, *IV-FGMM-ST-P*).

Le tableau 4.12 présente les temps de calcul mesurés (en secondes). Les méthodes sont classées dans l'ordre croissant du temps obtenu avec la séquence vidéo la plus longue.

On note que notre méthode se classe la 7<sup>e</sup> plus lente.

D'après les auteurs la méthode manquante *BMOG*, permet un calcul proche du temps réel.

Sur la figure 4.7, on observe qu'il y a de fortes disparités entre les méthodes : en particulier, certaines ACP robustes sont très coûteuses en temps, comme l'algorithme *LSADM*, et *EALM*. Mais d'autres ACP ro-

Méthode	50 images	150 images	250 images	450 images	5016 images
Durée séquence	1,67	5	8,33	15	167,200
FG	0,892	2,23	3,123	6,125	41,531
T2-FGMM-UM	0,552	1,531	2,335	4,965	41,531
T2-FGMM-UV	0,512	1,236	2,311	4,883	42,241
KDE-G	1,032	2,984	4,157	5,813	42,271
GMM-Z	0,423	1,134	2,045	4,164	48,542
GMM SG	0,434	1,248	2,124	4,234	49,021
T2-FGMM-VMRF	0,921	3,32	5,203	7,212	56,435
T2-FGMM-UMRF	0,892	3,301	4,85	7,983	57,243
FSI	1,01	3,012	4,02	9,103	80,105
FCI	1,01	4,102	6,13	11,12	86,341
IV-FGMM-P	2,721	9,064	13,357	23,71	232,971
IV-FGMM-T-P	2,903	10,174	14,265	25,66	251,340
FPCP	3,799	6,44	9,816	15,784	406,945
IV-FGMM	3,809	10,975	18,309	32,876	470,602
IV-FGMM-T	4,175	11,859	20,057	39,813	472,413
IV-FGMM-S-P	6,882	23,524	39,998	73,224	783,996
<b>IV-FGMM-ST-P</b>	<b>8,837</b>	<b>25,396</b>	<b>42,72</b>	<b>75,842</b>	<b>891,996</b>
ACP classique	4,999	6,86	10,337	18,88	1119,489
IV-FGMM-S	12,218	39,052	62,276	108,375	1252,930
<b>IV-FGMM-ST</b>	<b>12,839</b>	<b>42,993</b>	<b>64,137</b>	<b>112,512</b>	<b>1341,310</b>
RACP	4,097	12,695	28,936	109,483	> $t_{lim}$
IALM	12,108	36,494	69,74	171,047	> $t_{lim}$
MOG-ACP	43,306	75,692	144,462	196,163	> $t_{lim}$
ALM	73,024	116,278	111,886	268,9	> $t_{lim}$
LSADM	282,258	1074,795	1960,673	3956,424	> $t_{lim}$
EALM	243,345	1139,846	1621,497	5248,897	> $t_{lim}$

TABEAU 4.12 – Temps de calcul des différents méthodes ( $t_{lim} = 20\ 000s$ ).

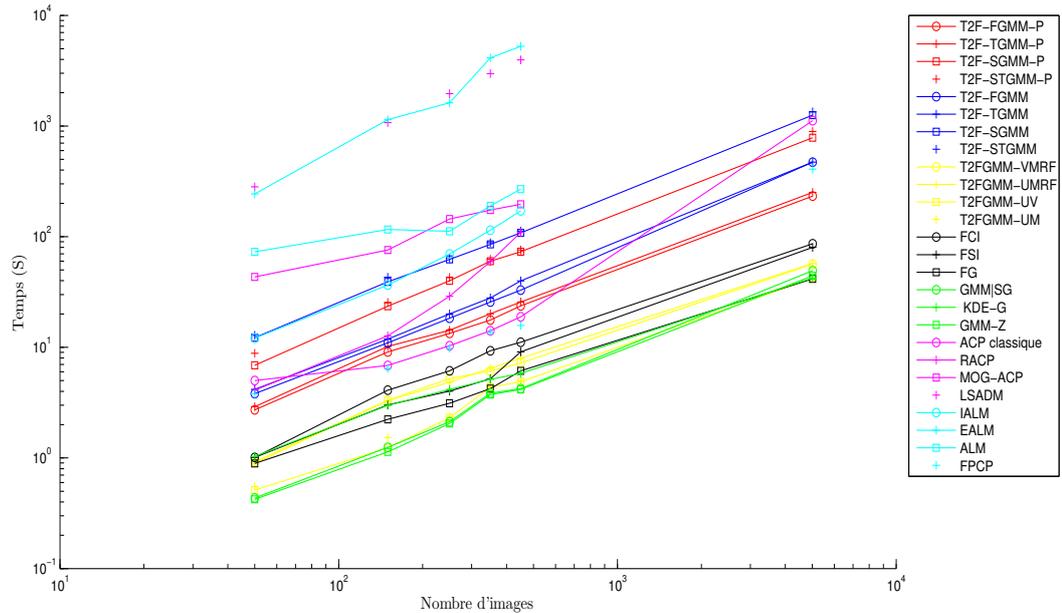


FIGURE 4.7 – Temps de calcul en fonction du nombre d’images de la séquence vidéo.

bustes ont des temps de calcul tout-à-fait satisfaisants, notamment *FPCP*. On note toutefois que sa croissance est plus forte que celle de notre méthode, lorsque le nombre d’images approche le maximum (5016) : sa complexité est donc d’un ordre supérieur.

Les méthodes *T2-FGMM* sont rapides, davantage que nos différentes variantes, qui ont pourtant des complexités similaires (pentes approximativement égales). Notre méthode n’est en effet pas économe sur les calculs, notamment ceux des intervalles. Et elle devient particulièrement coûteuse lorsqu’on intègre la fusion spatio-temporelle.

## 4.7 Conclusion du chapitre

Dans la première partie ce chapitre, nous avons montré l’importance de chaque étape de notre méthode, ainsi que son efficacité sur la vidéo *Canoé*, en dépit d’un fond dynamique perturbant.

Notre méthode fournit globalement de bons résultats, qui peuvent encore être améliorés en optimisant son jeu de paramètre pour *chaque* vidéo (c’est un usage largement répandu).

Dans la seconde partie de ce chapitre, nous avons comparé notre méthode à celles de la littérature, basées sur le flou, le modèle *GMM* ou l’*ACP*. Les résultats tendent à montrer que notre méthode donne d’au moins aussi bons résultats, avec des temps de calculs raisonnables, que la plupart de ces méthodes.

Il faut cependant noter les scores remarquables des méthodes *RMoG* et *BMOG*. Leur supériorité par rapport à notre méthode s’explique sans doute par la nature différente de l’information qu’elles extraient des images.

## 4.8 Références

- BAF, F. E., T. BOUWMANS et B. VACHON. 2008, «Fuzzy integral for moving object detection», dans *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, ISSN 1098-7584, p. 1729–1736, doi :10.1109/FUZZY.2008.4630604. 84
- DARWICH, A., P.-A. HÉBERT, Y. MOHANNA et A. BIGAND. 2015, «Analyse en composantes principales, application a la modélisation de l'arrière-plan», dans *JFL3 Liban*. 84
- EL BAF, F., T. BOUWMANS et B. VACHON. 2008, *Type-2 Fuzzy Mixture of Gaussians Model : Application to Background Modeling*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-89639-5, p. 772–781, doi :10.1007/978-3-540-89639-5\_74. URL [https://doi.org/10.1007/978-3-540-89639-5\\_74](https://doi.org/10.1007/978-3-540-89639-5_74). 84
- ELGAMMAL, A., D. HARWOOD et L. DAVIS. 2000, *Non-parametric Model for Background Subtraction*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-45053-5, p. 751–767, doi :10.1007/3-540-45053-X\_48. URL [https://doi.org/10.1007/3-540-45053-X\\_48](https://doi.org/10.1007/3-540-45053-X_48). 84
- GOLDFARB, D., S. MA et K. SCHEINBERG. 2013, «Fast alternating linearization methods for minimizing the sum of two convex functions», *Mathematical Programming*, p. 1–34. 84
- LIN, Z., M. CHEN et Y. MA. 2010, «The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices», *arXiv preprint arXiv :1009.5055*. 84
- MARTINS, I., P. CARVALHO, L. CORTE-REAL et J. L. ALBA-CASTRO. 2017, *BMOG : Boosted Gaussian Mixture Model with Controlled Complexity*, Springer International Publishing, Cham, ISBN 978-3-319-58838-4, p. 50–57, doi :10.1007/978-3-319-58838-4\_6. URL [https://doi.org/10.1007/978-3-319-58838-4\\_6](https://doi.org/10.1007/978-3-319-58838-4_6). 84
- RODRÍGUEZ, P. et B. WOHLBERG. 2013, «Fast principal component pursuit via alternating minimization», dans *2013 IEEE International Conference on Image Processing*, ISSN 1522-4880, p. 69–73, doi :10.1109/ICIP.2013.6738015. 84
- SIGARI, M. H., N. MOZAYANI et H. POURREZA. 2008, «Fuzzy running average and fuzzy background subtraction : concepts and application», *International Journal of Computer Science and Network Security*, vol. 8, n° 2, p. 138–143. 84
- SOBRAL, A. 2013, «Bgslibrary : An opencv c++ background subtraction library», dans *IX Workshop de Visão Computacional (WVC'2013)*, Rio de Janeiro, Brazil. URL <https://github.com/andrewssobral/bgslibrary>. 76, 84
- SOBRAL, A., T. BOUWMANS et E.-H. ZAHZAH. 2015, «Lrslibrary : Low-rank and sparse tools for background modeling and subtraction in videos», dans *Robust Low-Rank and Sparse Matrix Decomposition : Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group. 76, 84

- STAUFFER, C. et W. E. L. GRIMSON. 1999, «Adaptive background mixture models for real-time tracking», dans *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 2, ISSN 1063-6919, p. 252 Vol. 2, doi :10.1109/CVPR.1999.784637. 84
- TANG, G. et A. NEHORAI. 2011, «Robust principal component analysis based on low-rank and block-sparse matrix decomposition», dans *2011 45th Annual Conference on Information Sciences and Systems*, p. 1–5, doi :10.1109/CISS.2011.5766144. 84
- DE LA TORRE, F. et M. J. BLACK. 2001, «Robust principal component analysis for computer vision», dans *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, IEEE, p. 362–369. 84
- VARADARAJAN, S., P. MILLER et H. ZHOU. 2013, «Spatial mixture of gaussians for dynamic background modelling», dans *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, p. 63–68, doi :10.1109/AVSS.2013.6636617. 84
- WANG, Y., P.-M. JODOIN, F. PORIKLI, J. KONRAD, Y. BENEZETH et P. ISHWAR. 2014, «Cdnets 2014 : An expanded change detection benchmark dataset», dans *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW '14*, IEEE Computer Society, Washington, DC, USA, ISBN 978-1-4799-4308-1, p. 393–400, doi :10.1109/CVPRW.2014.126. URL <http://dx.doi.org/10.1109/CVPRW.2014.126>. 82, 83
- ZHANG, H. et D. XU. 2006, *Fusing Color and Texture Features for Background Model*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-45917-0, p. 887–893, doi :10.1007/11881599\_110. URL [https://doi.org/10.1007/11881599\\_110](https://doi.org/10.1007/11881599_110). 84
- ZHAO, Q., D. MENG, Z. XU, W. ZUO et L. ZHANG. 2014, «Robust principal component analysis with complex noise», dans *International Conference on Machine Learning*, p. 55–63. 84
- ZHAO, Z., T. BOUWMANS, X. ZHANG et Y. FANG. 2012, *A Fuzzy Background Modeling Approach for Motion Detection in Dynamic Backgrounds*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-35286-7, p. 177–185, doi :10.1007/978-3-642-35286-7\_23. URL [https://doi.org/10.1007/978-3-642-35286-7\\_23](https://doi.org/10.1007/978-3-642-35286-7_23). 84
- ZIVKOVIC, Z. 2004, «Improved adaptive gaussian mixture model for background subtraction», dans *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, ISSN 1051-4651, p. 28–31 Vol.2, doi :10.1109/ICPR.2004.1333992. 84



# Conclusion générale et perspectives

## Conclusion générale

De nombreux travaux de recherche ont été récemment consacrés à la soustraction du fond, afin de développer des méthodes robustes aux changements dynamiques d'arrière-plans, à la variation de luminosité, et autres perturbations. Qu'elles soient supervisées ou non-supervisées, la plupart utilisent des modèles précis, dans lesquels l'incertitude sur les paramètres ou sur les données peut occasionner des erreurs de classement des pixels.

Dans le premier chapitre, nous avons rappelé les méthodes classiques de soustraction du fond, ainsi qu'un ensemble de méthodes basées sur l'*ACP*, classique ou robustes. Parmi toutes ces méthodes, nous avons insisté sur la méthode de Stauffer, qui malgré son âge, reste intéressante pour ses performances et sa rapidité d'exécution.

Dans le deuxième chapitre, nous avons présenté les différentes classes d'ensembles flous, dont ceux de type-1, de type-2, les nombres flous intuitifs, pythagoriciens, ainsi que les *IVFS*. Et nous avons souligné l'importance de la modélisation floue pour gérer l'incertitude, notamment dans son application au traitement des vidéos. Enfin, nous avons mis en valeur la méthode *T2FGMM*, basée sur un modèle de mélange gaussien flou de type *IVFS*.

Dans le chapitre 3, nous avons proposé une méthode originale de soustraction du fond non-supervisée, basée sur un mélange gaussien flou de type-2, particulièrement robuste aux changements dynamiques du fond. La méthode procède par pixel, en calculant puis agrégeant les réponses de différents classifieurs flous complémentaires, exploitant chacun une part d'information du pixel : sa couleur, les décisions de son voisinage, et celle de son (ou de ses) pixels antécédents dans les images précédentes, identifiés par un calcul de flot optique.

Dans le dernier chapitre, nous avons d'abord effectué une validation complète et critique des différentes composantes de la méthode, sur une vidéo de référence. Nous avons ainsi confirmé la pertinence de l'utilisation du flou, et des différents classifieurs (spatial, temporel), même si le classifieur temporel mériterait sans doute des améliorations.

Puis nous avons comparé la méthode à des algorithmes performants de la littérature. Pour cette comparaison nous avons utilisé le catalogue d'arrière-plans dynamiques de la base de données de référence **CD-NET 2014**. Malgré des temps de calcul assez importants, que l'on peut heureusement minimiser grâce à la parallélisation, la méthode s'est révélée très performante par rapport à la plupart des méthodes testées. Cela atteste de sa capacité à traiter les vidéos à changements dynamiques de fond.

## Perspectives

De nombreuses perspectives peuvent émerger de ce travail.

Tout d’abord, sans remettre en cause le principe de la méthode, il serait intéressant de simplifier les calculs, de manière à réduire les temps d’exécution. En effet, la complexité de la méthode est tout-à-fait raisonnable. Le seul problème provient de l’importante quantité de calculs à réaliser, pour chaque pixel, de chaque image : en particulier les évaluations des fonctions d’appartenance imprécises pour le calcul des intervalles et les fusion des intervalles. Pour travailler avec des images de grande résolution, il est donc nécessaire d’allier les mécanismes de parallélisation à une approximation des calculs.

Toujours sans remettre en cause l’algorithme, on pourrait enrichir nos classifieurs flous, notamment avec des informations de textures, ou de couleurs, dans d’autres espaces que celui *RGB*. La fusion temporelle mériterait également d’être améliorée, car sa contribution actuelle aux performances de la méthode est faible.

En ce qui concerne la prise en charge des incertitudes dans le modèle de fond, nous nous sommes tournés - pour leur simplicité - vers les ensembles flous de type *IVFS*. On pourrait leur préférer le formalisme plus riche des nombres pythagoriciens, et voir si la méthode peut alors gagner en performance.

Sur le modèle de fond lui-même, actuellement un mélange de gaussiennes, il y a également des perspectives d’amélioration : il serait judicieux de l’initialiser avec un *EM* flou, plutôt que d’insérer le flou a posteriori de manière artificielle. On pourrait également envisager de le construire avec une version floue de la méthode de Parzen. D’autre part, au fur et à mesure de l’accumulation des images, il serait pertinent de réguler la quantité de « flou », plutôt que de la fixer a priori pour toute la vidéo.

Du point de vue de l’intérêt applicatif, une poursuite logique de ce travail serait d’intégrer le suivi des objets, avec notamment une discrimination supplémentaire *Objet connu* contre *Objet inconnu/nouveau*. Cela permettrait par exemple de faire du suivi de personnes en ville, et de participer ainsi au concept moderne de « *smart city* ».

Finalement, les techniques récentes d’apprentissage profond (« deep learning ») constituent une autre piste de recherche prometteuse. Récemment, BRAHAM et VAN DROOGENBROECK [2016]; WANG et collab. [2017] ont commencé à les employer pour la modélisation du fond, avec succès, pour toutes les catégories de vidéos. On pourrait par exemple envisager d’intégrer une gestion des incertitudes, par ensembles flous, dans ces techniques.

## Références

- BRAHAM, M. et M. VAN DROOGENBROECK. 2016, «Deep background subtraction with scene-specific convolutional neural networks», dans *Systems, Signals and Image Processing (IWSSIP), 2016 International Conference on*, IEEE, p. 1–4. [94](#)
- WANG, Y., Z. LUO et P.-M. JODOIN. 2017, «Interactive deep learning method for segmenting moving objects», *Pattern Recognition Letters*, vol. 96, p. 66–75. [94](#)

# Annexe A

## Annexe

### A.1 Tableaux de résultats pour toutes les méthodes basées sur le *GMM*

Méthode	TP%	FP%	FN%	TN%	Spécificité	PBC	Rappel	Précision	F-mesure
IV-FGMM-ST	0,794	0,555	0,638	998,013	0,999	0,119	0,554	0,589	0,571
GMM-Z	4,393	1,090	1,879	992,638	0,999	0,297	0,700	0,801	0,747
KDE-G	4,124	2,649	2,148	991,079	0,997	0,48	0,658	0,609	0,632
GMM SG	4,755	2,024	1,517	991,70	0,998	0,354	0,758	0,701	0,729
<b>RMoG</b>	5,137	0,934	1,135	992,794	0,999	0,207	0,819	0,846	<b>0,832</b>
<b>BMOG</b>	4,671	0,194	1,601	993,534	1	0,179	0,745	0,960	<b>0,839</b>
T2-FGMM-UM	0,017	0,027	0,565	999,391	1	0,059	0,029	0,384	0,054
T2-FGMM-UV	0,049	8,273	0,533	991,145	0,992	0,881	0,085	0,006	0,011
T2-FGMM-UMRF	0,019	0,041	0,563	999,377	1	0,060	0,032	0,313	0,058
T2-FGMM-VMRF	0,046	8,710	0,536	990,707	0,991	0,925	0,079	0,005	0,010
FSI	1,457	1,106	3,663	993,774	0,999	0,477	0,285	0,568	0,379
FCI	1,897	1,816	3,223	993,064	0,998	0,504	0,370	0,511	0,429
FG	5,003	520,665	0,117	474,215	0,477	52,078	0,977	0,010	0,019

TABLEAU A.1 – Vidéo boats

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	20,353	0,374	3,135	976,138	1	0,351	0,867	0,982	<b>0,921</b>
GMM-Z	30,216	2,647	5,197	961,940	0,997	0,784	0,853	0,919	0,885
KDE-G	29,444	1,891	5,968	962,696	0,998	0,786	0,831	0,940	0,882
GMM SG	30,664	3,476	4,749	961,111	0,996	0,822	0,866	0,898	0,882
<b>RMoG</b>	31,928	0,905	3,485	963,683	0,999	0,439	0,902	0,972	<b>0,936</b>
<b>BMOG</b>	33,460	1,533	1,953	963,054	0,998	0,349	0,945	0,956	<b>0,950</b>
T2-FGMM-UM	0,227	0,172	2,885	996,716	1	0,306	0,073	0,569	0,129
T2-FGMM-UV	0,375	0,116	2,737	996,771	1	0,285	0,121	0,763	0,208
T2-FGMM-UMRF	0,382	0,430	2,730	996,458	1	0,316	0,123	0,471	0,195
T2-FGMM-VMRF	0,365	0,114	2,747	996,773	1	0,286	0,117	0,762	0,203
FSI	21,687	0,803	8,707	968,803	0,999	0,951	0,714	0,964	0,820
FCI	23,066	1,673	7,327	967,933	0,998	0,900	0,759	0,932	0,837
FG	29,992	396,136	0,401	573,469	0,591	39,654	0,987	0,070	0,131

TABLEAU A.2 – Vidéo Canoé.

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	7,215	3,136	5,084	984,565	0,997	0,822	0,587	0,697	<b>0,637</b>
GMM-Z	15,165	38,678	2,550	943,606	0,961	4,123	0,856	0,282	0,424
KDE-G	15,450	66,936	2,266	915,348	0,932	6,92	0,872	0,188	0,309
GMM SG	15,657	38,474	2,059	943,811	0,961	4,053	0,884	0,289	0,436
<b>RMoG</b>	12,671	7,264	5,045	975,021	0,993	1,231	0,715	0,636	<b>0,673</b>
<b>BMOG</b>	17,367	15,147	0,349	967,137	0,985	1,55	0,98	0,534	<b>0,692</b>
T2-FGMM-UM	0,132	3,495	0,295	996,079	0,997	0,379	0,309	0,036	0,065
T2-FGMM-UV	0,130	16,624	0,295	982,951	0,983	1,692	0,306	0,008	0,015
T2-FGMM-UMRF	0,132	3,495	0,294	996,079	0,997	0,379	0,31	0,036	0,065
T2-FGMM-VMRF	0,130	16,624	0,295	982,951	0,983	1,692	0,306	0,008	0,015
FSI	9,545	57,959	7,578	924,918	0,941	6,554	0,557	0,141	0,226
FCI	10,443	69,534	6,680	913,343	0,929	7,621	0,61	0,131	0,215
FG	16,222	414,105	0,901	568,772	0,579	41,501	0,947	0,038	0,073

TABLEAU A.3 – Vidéo Fall.

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	6,792	0,097	1,564	991,547	1	0,166	0,813	0,986	<b>0,891</b>
GMM-Z	10,820	0,733	2,578	985,869	0,999	0,331	0,808	0,937	0,867
KDE-G	10,723	1,875	2,675	984,726	0,998	0,455	0,800	0,851	0,825
GMM SG	11,113	0,979	2,286	985,623	0,999	0,326	0,829	0,919	0,872
<b>RMoG</b>	11,314	0,400	2,084	986,202	1	0,248	0,844	0,966	<b>0,901</b>
<b>BMOG</b>	13,249	0,890	0,150	985,712	0,999	0,104	0,989	0,937	<b>0,962</b>
T2-FGMM-UM	0,441	0,384	0,834	998,341	1	0,122	0,346	0,534	0,420
T2-FGMM-UV	0,567	1,112	0,708	997,613	0,999	0,182	0,445	0,338	0,384
T2-FGMM-UMRF	0,472	0,673	0,803	998,052	0,999	0,148	0,370	0,412	0,390
T2-FGMM-VMRF	0,539	1,070	0,735	997,655	0,999	0,181	0,423	0,335	0,374
FSI	7,050	11,030	4,553	977,367	0,989	1,558	0,608	0,390	0,475
FCI	7,458	17,869	4,145	970,529	0,982	2,201	0,643	0,294	0,404
FG	11,243	219,238	0,360	769,159	0,778	21,96	0,969	0,049	0,093

TABLEAU A.4 – Vidéo Overpass.

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
IV-FGMM-ST	0,115	2,652	0,159	997,074	0,997	0,281	0,420	0,042	0,076
GMM-Z	0,623	13,838	0,207	985,332	0,986	1,404	0,751	0,043	0,081
KDE-G	0,658	10,991	0,172	988,179	0,989	1,116	0,793	0,057	0,106
GMM SG	0,662	15,849	0,168	983,321	0,984	1,602	0,797	0,040	0,076
<b>RMoG</b>	0,455	3,194	0,375	995,976	0,997	0,357	0,548	0,125	<b>0,203</b>
<b>BMOG</b>	0,655	1,951	0,175	997,219	0,998	0,213	0,789	0,251	<b>0,381</b>
T2-FGMM-UM	0,040	2,343	0,133	997,484	0,998	0,248	0,230	0,017	0,031
T2-FGMM-UV	0,047	1,428	0,126	998,399	0,999	0,155	0,270	0,032	0,057
T2-FGMM-UMRF	0,042	3,767	0,131	996,060	0,996	0,390	0,245	0,011	0,021
T2-FGMM-VMRF	0,042	1,408	0,131	998,419	0,999	0,154	0,241	0,029	0,051
FSI	0,264	6,911	0,325	992,500	0,993	0,724	0,449	0,037	0,068
FCI	0,291	10,453	0,298	988,958	0,990	1,075	0,494	0,027	0,051
FG	0,524	188,919	0,065	810,492	0,811	18,898	0,889	0,003	0,006

TABLEAU A.5 – Vidéo Fontaine1.

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
IV-FGMM-ST	1,293	0,027	0,138	998,542	1	0,017	0,903	0,980	<b>0,940</b>
GMM-Z	1,815	0,618	0,336	997,230	0,999	0,095	0,844	0,746	0,792
KDE-G	1,834	0,471	0,317	997,377	1	0,079	0,853	0,796	0,823
GMM SG	1,875	0,641	0,276	997,207	0,999	0,092	0,872	0,745	0,803
<b>RMoG</b>	1,952	0,406	0,200	997,443	1	0,061	0,907	0,828	<b>0,866</b>
<b>BMOG</b>	2,055	0,203	0,096	997,646	1	0,030	0,956	0,910	<b>0,932</b>
T2-FGMM-UM	0,112	0,030	0,207	999,652	1	0,024	0,351	0,790	0,486
T2-FGMM-UV	0,142	0,074	0,177	999,607	1	0,025	0,445	0,656	0,530
T2-FGMM-UMRF	0,124	0,075	0,195	999,607	1	0,027	0,389	0,624	0,479
T2-FGMM-VMRF	0,132	0,074	0,187	999,608	1	0,026	0,413	0,641	0,502
FSI	0,644	0,244	1,000	998,111	1	0,124	0,392	0,725	0,509
FCI	0,747	0,410	0,897	997,946	1	0,131	0,454	0,646	0,533
FG	1,603	196,958	0,041	801,398	0,803	19,700	0,975	0,008	0,016

TABLEAU A.6 – Vidéo Fontaine2.

## A.2 Tableaux de résultats pour toutes les méthodes basées sur l'ACP

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	6,586	0,280	2,329	990,804	1	0,261	0,739	0,959	<b>0,835</b>
ACP-classique	4,210	4,254	9,202	982,334	0,996	1,346	0,314	0,497	0,385
RACP	5,657	5,035	7,755	981,552	0,995	1,279	0,422	0,529	0,469
LSADM	5,019	4,060	8,393	982,528	0,996	1,245	0,374	0,553	0,446
<b>ALM</b>	10,932	0,151	2,480	986,437	1	0,263	0,815	0,986	<b>0,893</b>
EALM	5,096	3,119	8,316	983,469	0,997	1,143	0,380	0,620	0,471
IALM	5,101	3,114	8,321	983,465	0,997	1,143	0,380	0,621	0,472
<b>FPCP</b>	10,933	0,151	2,479	986,436	1	0,263	0,815	0,986	<b>0,893</b>
MOG-ACP	6,481	12,464	6,931	974,123	0,987	1,940	0,483	0,342	0,401

TABLEAU A.7 – Vidéo Canoé

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	8,538	2,272	0,105	989,086	0,998	0,238	0,988	0,790	<b>0,878</b>
ACP-classique	6,856	4,530	2,417	986,197	0,995	0,695	0,739	0,602	0,664
RACP	7,005	1,537	2,268	989,190	0,998	0,380	0,755	0,820	0,786
<b>LSADM</b>	6,615	0,216	2,658	990,511	1	0,287	0,713	0,968	<b>0,822</b>
ALM	7,340	0,420	1,933	990,307	1	0,235	0,792	0,946	0,862
EALM	6,743	0,642	2,530	990,085	0,999	0,317	0,727	0,913	0,810
IALM	6,708	0,617	2,565	990,111	0,999	0,318	0,723	0,916	0,808
<b>FPCP</b>	7,340	0,421	1,933	990,306	1	0,235	0,792	0,946	<b>0,862</b>
MOG-ACP	7,386	0,612	1,887	990,115	0,999	0,250	0,796	0,923	0,855

TABLEAU A.8 – Vidéo Overpass

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	0,055	0,305	0,142	999,498	1	0,045	0,279	0,153	<b>0,197</b>
ACP-classique	0,248	11,781	0,119	987,851	0,988	1,190	0,675	0,021	0,040
RACP	0,232	12,193	0,136	987,439	0,988	1,233	0,631	0,019	0,036
LSADM	0,148	5,096	0,220	994,536	0,995	0,532	0,403	0,028	0,053
ALM	0,230	15,067	0,138	984,565	0,985	1,520	0,626	0,015	0,029
EALM	0,250	9,972	0,118	989,660	0,990	1,009	0,679	0,024	0,047
IALM	0,249	9,798	0,119	989,835	9,90	9,92	6,76	0,25	0,48
FPCP	0,228	15,070	0,135	984,567	0,985	1,521	0,627	0,015	0,029
MOG-ACP	0,234	12,111	0,134	987,521	0,988	1,225	0,635	0,019	0,037

TABLEAU A.9 – Vidéo Fontaine1

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	0,715	0,061	0,081	999,143	1	0,014	0,898	0,921	<b>0,909</b>
ACP-classique	0,962	0,903	0,103	998,033	0,999	0,101	0,904	0,516	0,657
RACP	0,969	0,794	0,096	998,142	0,999	0,089	0,910	0,550	0,685
LSADM	0,826	0,333	0,239	998,602	1	0,057	0,776	0,713	0,743
ALM	0,935	1,160	0,130	997,775	0,999	0,129	0,878	0,446	0,592
EALM	0,938	1,157	0,128	997,777	0,999	0,129	0,880	0,448	0,593
IALM	0,983	0,900	0,082	998,036	0,999	0,098	0,923	0,522	0,667
FPCP	0,935	1,161	0,130	997,774	0,999	0,129	0,878	0,446	0,592
MOG-ACP	0,969	0,830	0,096	998,105	0,999	0,093	0,910	0,538	0,677

TABLEAU A.10 – Vidéo Fontaine2

Méthode	TP‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
IV-FGMM-ST	0,237	0,013	2,807	996,943	1	0,282	0,078	0,947	0,144
ACP-classique	1,703	0,117	1,650	996,530	1	0,177	0,508	0,936	0,658
RACP	1,704	0,117	1,650	996,530	1	0,177	0,508	0,936	0,659
LSADM	1,853	0,083	1,501	996,563	1	0,158	0,552	0,957	0,701
ALM	1,852	0,084	1,500	996,564	1	0,158	0,552	0,957	0,700
<b>EALM</b>	1,853	0,083	1,501	996,563	1	0,158	0,553	0,957	0,701
IALM	1,944	0,104	1,409	996,543	1	0,151	0,580	0,949	<b>0,720</b>
FPCP	1,853	0,084	1,500	996,564	1	0,158	0,553	0,957	0,701
<b>MOG-ACP</b>	1,989	0,270	1,364	996,377	1	0,163	0,593	0,880	<b>0,709</b>

TABLEAU A.11 – Vidéo boats

Méthode	TP ‰	FP‰	FN‰	TN‰	Spécificité	PBC	Rappel	Précision	F-mesure
<b>IV-FGMM-ST</b>	9,786	0,092	1,425	988,697	1	0,152	0,873	0,991	<b>0,928</b>
ACP-classique	6,537	4,350	8,527	980,585	0,996	1,288	0,434	0,600	0,504
RACP	6,440	5,232	8,624	979,704	0,995	1,386	0,428	0,552	0,482
LSADM	8,354	8,403	6,711	976,531	0,991	1,511	0,555	0,499	0,525
ALM	8,354	8,404	6,710	976,532	0,991	1,511	0,555	0,499	0,525
EALM	8,355	8,403	6,711	976,531	0,991	1,511	0,555	0,499	0,525
IALM	9,341	2,945	5,723	981,990	0,997	0,867	0,620	0,760	0,683
FPCP	8,354	8,406	6,710	976,530	0,991	1,512	0,555	0,498	0,525
MOG-ACP	8,724	4,824	6,341	980,111	0,995	1,117	0,579	0,644	0,610

TABLEAU A.12 – Vidéo fall.

