

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de Guelma

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

**Mémoire de fin d'étude Master 2**



Département d'Informatique

Spécialité : Ingénierie des médias

Thème :

---

---

**Détection, Segmentation et Suivi d'un individu en mouvement  
dans une vidéo de surveillance**

---

---

Présenté par : Bouharkat Asma

Belhaoues Sarah

Sous la direction de :

Brahim Farou

Juin 2011

To Remove Please Registered With <http://www.pdf-converter-software.com>

PDF CONVERTER SOFTWARE

## *Remerciements*

Nous remercions vivement tous ceux qui ont participé de loin ou de près par une idée ou une suggestion, par un moment de soutien moral et de concours intellectuel à l'élaboration de projet.

Nos remerciements, nous les semons à tout vent en direction de notre honorable jury et avec cœur ouvert à notre cher encadreur.

## • Résumé

Au début des années 2000, les caméras font leur apparition en nombre important dans de nombreuses villes européennes. La télésurveillance devient une technique de plus en plus utilisée dans plusieurs domaines, entre autres, la surveillance à distance d'un lieu, public ou privé, de machines ou d'individus, généralement pour des raisons de sécurité que se soit dans le cadre de la sécurité routière, le cadre de la prévention de la délinquance, le cadre de la télé-médecine, pour la surveillance à distance des enfants et des personnes vulnérables et encore pour la surveillance de lieux sensibles (banques, centrales nucléaires, etc.) et d'habitations, afin de prévenir les intrusions, les cambriolages et les actes de vandalisme. [FSHS, 09]

Le présent mémoire est structuré en trois chapitres : le premier chapitre est consacré à un état de l'art des images et de la vidéo, le deuxième chapitre est consacré à la conception du système, et le dernier pour les tests et résultats. L'objectif consiste à réaliser un système qui permettra une fois achevée la détection, la segmentation en partie et le suivi des activités d'un individu dans une vidéo de surveillance extraite à partir d'une caméra fixe.

**Mots-clés :** vision, traitement d'image, segmentation, détection des événements, comportement d'un individu.

www.pdf-converter-software.com

# • Sommaire

<b>Introduction générale</b> .....	7
<i>Chapitre I : Traitement d'image et de la vidéo</i>	
1. Introduction .....	9
2. La vidéo.....	9
2.1 C'est quoi une vidéo ?.....	9
2.2 Domaine d'utilisation de la vidéo.....	9
2.3 Caractéristiques de la vidéo .....	11
2.3.1. Nombre d'images par seconde .....	11
2.3.2. Entrelacement .....	11
2.3.3. Résolution de l'affichage .....	12
2.3.4 .Allongement .....	12
2.3.5. Couleur espace et de bits par pixel .....	13
2.3.6. Qualité vidéo .....	13
2.3.7. Le débit .....	14
3. Les images.....	15
3.1 Qu'est-ce qu'une image.....	15
3.2 Les caractéristiques d'une image.....	15
3.2.1 Pixels.....	15
3.2.2 La dimension.....	15
3.2.3 La résolution.....	16
3.2.4 Le codage de la couleur.....	16
<b>3.2.5 Luminance</b> .....	16
3.2.6 Contraste.....	16
3.2.7 Format.....	17

PDF TO PDF CONVERTER SOFTWARE

- 4. Séparation avant plan \arrière plan.....17
  - 4.1 Avant plan.....17
  - 4.2 Arrière-plan.....17
  - 4.3 Séparation avant-plan/arrière-plan par couleur.....18
  - 4.4 Séparation avant-plan/arrière-plan par contours.....18
  - 4.5 Séparation avant-plan/arrière-plan par position.....19
  - 4.6 Séparation avant-plan/arrière-plan par ondelettes.....20
- 5. Les méthodes de détection et de classification.....20
  - 5.1 Approche globale de l’image (pixellaire).....20
    - 5.1.1 Les méthodes à seuils.....20
    - 5.1.2 La classification (clustering).....21
  - 5.2 Approche région.....22
    - 5.2.1 La croissance de région.....22
    - 5.2.2 Les méthodes de décomposition et fusion .....23
  - 5.3 Approche contour.....24
    - 5.3.1 Les méthodes de calculs de gradients discrets.....24
    - 5.3.2 Les détecteurs de contour.....24
    - 5.3.3 Les modèles déformables.....24
  - 5.4 Approche coopérative.....25
    - 5.4.1 Par fusion de résultats obtenus avec diverses méthodes.....25
    - 5.4.2 Par utilisations successives des approches contour et région.....25
    - 5.4.3 Par utilisation simultanée de méthodes basées contour et région.....25
- 6. Les méthodes d’extraction fond/forme.....26
  - 6.1 Mixture de gaussienne.....26
  - 6.2 Codebook2 layers.....27
  - 6.3 Vumètre.....29
- 7. Conclusion.....31

PDF CONVERTER SOFTWARE

## **Chapitre II : Conception du système**

1. Introduction .....	32
2. Problématique .....	32
3. Architecture du système .....	33
4. Réalisation de l'architecture .....	34
4.1 Phase 1.....	34
4.2 Phase 2.....	34
4.2.1 Lecture vidéo.....	34
4.2.2 Fragmenter vidéo.....	34
4.2.3 Image fragmenté .....	35
4.3 Phase 3.....	35
4.3.1 Niveau de gris.....	35
4.3.2 Gaussienne multiple.....	36
4.3.2.a Apprentissage.....	36
4.3.2.b Extraction de l'avant plan.....	36
4.3.2.c Poste traitement .....	38
1. Dilatation automatique.....	38
2 .Elimination des points isolés.....	41
4.3.3 Suivi d'objet.....	41
4.4 Phase 4.....	43
5. Conclusion.....	44

## **Chapitre III : Tests et résultats**

1. Introduction.....	45
2. Base de vidéo.....	45
2.1 Vidéo de type milieu naturel ouvert.....	45
2.2 Vidéo dans un milieu fermé.....	46

2.3 Vidéo dans un milieu type administratif.....	47
3. Les tests et résultats.....	48
3.1 Résultats de l'algorithme des gaussiennes multiples .....	48
3.2 Résultats des algorithmes de poste traitement.....	50
3.2.1 Dilatation automatique.....	50
3.2.2 Elimination des points isolés.....	51
3.3 Résultats de l'algorithme de suivi d'objets.....	51
3.4 Cas où la dilatation n'est pas un succès.....	52
4. Conclusion.....	54
<b>Conclusion générale.....</b>	<b>55</b>
Bibliographie.....	56





www.pdf-converter-software.com

## • Liste des tableaux

Tableau 1.1 : Les principaux formats d'une image .....17

## • Introduction générale

Cette menace qui a toujours été présente n'a jamais vraiment créé un sentiment d'insécurité, mais les attentats du 11 septembre 2001 ont changé la donnée. Les gens ont pris conscience que personne n'était intouchable. Toutefois la mise en place de la vidéosurveillance ne peut s'expliquer uniquement par l'insécurité grandissante ou la protection des biens. Certaines autres raisons moins connues du grand public existent également. La mise en place de la vidéosurveillance permet une amélioration de la gestion des incidents ainsi qu'une augmentation de l'efficacité et de la rapidité d'intervention. Par exemple, dans la prévention du suicide ou encore lors d'accidents qui pourraient survenir sur la voie publique. Elle permet ainsi indirectement, de maintenir les primes d'assurances à un niveau raisonnable. La surveillance des axes routiers sert à informer en temps réel les automobilistes sur les conditions du trafic.

Quelques affaires de crimes ont été résolues grâce aux enregistrements fournis par les caméras de surveillance. La question de savoir si la vidéosurveillance prémunit ou réduit les crimes n'a pas pu être montrée par les études indépendantes qui furent conduites que ce soit en Algérie ou à l'étranger.

Le gouvernement britannique a jugé de son côté que les effets bénéfiques n'étaient pas possibles à évaluer, bien que Scotland Yard ait affirmé, en 2008, que la vidéosurveillance à Londres, qui compte 500 000 caméras, n'avait permis d'élucider que 3 % des vols dans la rue. [FSHS, 09]

L'objectif du travail présenté dans ce mémoire consiste à implémenter une application pour la détection, la segmentation et le suivi d'un individu en mouvement dans son environnement vital. Pour cela nous avons partagé notre travail en trois grandes parties organisées comme suit :

Dans la première partie nous avons présenté un aperçu général sur la vidéo, leurs caractéristiques et leur domaine d'application, la même chose avec les images, ensuite nous avons résumé les différentes méthodes utilisés pour traiter les images constituant la vidéo. La deuxième partie est consacrée pour expliquer la problématique posée dans notre travail et détailler l'architecture proposée pour la résoudre.

To Remove Please Registered With <http://www.pdf-converter-software.com>

Et enfin, dans la troisième partie, nous avons réalisé des tests et présenté une étude sur les différents résultats obtenus.

PDF CONVERTER SOFTWARE

## 1. Introduction :

Le but de l'analyse d'image réside dans la description (forme, couleur, texture) et/ou la quantification (nombre, densité) des divers objets qui composent l'image étudiée. L'analyse trouve sa place dans de nombreux domaines tels que la robotique, les sciences de la vie ou encore la sécurité. [AP, 07]

Depuis la fin des années 90, la numérisation des contenus et la progression de puissance des ordinateurs ont rendu possible le traitement en temps réel des images de la vidéo pour en extraire des interprétations (que voit-on à l'image, que se passe-t-il, qui va où, etc). D'abord effectués en noir et blanc, puis en couleur, ces traitements ont commencé à sortir des laboratoires de recherche dans cette période, et à constituer des solutions exploitables, d'abord pour la surveillance routière, puis pour la surveillance de personnes et d'objets, et plus récemment pour la biométrie faciale. [FSHS, 09]

Dans la suite de cette partie, nous présentons un état de l'art des différentes approches existantes en domaine de traitement de l'image et de la vidéo.

## 2 .La vidéo :

### 2.1 C'est quoi une vidéo ?

La vidéo regroupe l'ensemble des techniques, technologie, permettant l'enregistrement ainsi que la restitution d'images animées, accompagnées ou non de son, sur un support adapté à l'électronique et non de type photochimique. Le mot vidéo vient du latin vidéo qui signifie « je vois ». C'est l'apocope de vidéophonie ou vidéogramme.

Un flux vidéo est composé d'une succession d'images, 25 par seconde en Europe (30 par seconde aux USA), composant l'illusion du mouvement. Chaque image est décomposée en lignes horizontales, chaque ligne pouvant être considérée comme une succession de points. La lecture et la restitution d'une image s'effectue donc séquentiellement ligne par ligne comme un texte écrit : de gauche à droite puis de haut en bas. [S1]

### 2.2 Domaine d'utilisation de la vidéo :

Il est tout à fait possible d'utiliser la vidéo dans plusieurs domaines, mais la vidéosurveillance considère comme le plus célèbre domaine d'utilisation de la vidéo avec un champ d'application illimité, on peut recompter : l'industrie, le transport privé ou collectif, le commerce et la distribution, les administrations et les services publics, la santé, les lieux public, l'enseignement, les banques....etc.

La vidéosurveillance consiste à placer des caméras de surveillance dans un lieu public ou privé pour visualiser en un endroit centralisé tous les flux de personnes au sein d'un lieu ouvert au public et prévenir vols, agressions et mouvements de foule.

Un autre domaine où on peut trouver la vidéo utilisée largement c'est la visioconférence ou vidéoconférence, qui peut être présente dans plusieurs secteurs comme :

□ **Le secteur privé :** grande entreprise, petite et moyenne entreprise, optimisation de la communication collaborative (filiales, partenaires, salariés,...) avec la mise en place d'une infrastructure de visioconférence.

□ **Le secteur public :**  
Collectivités locales (justice, administration,...). [S2]

□ **Education :**  
Le télé-enseignement (e-learning = apprentissage via internet) est une application adaptée notamment à l'enseignement spécialisé, soit dans les universités, permettant de bénéficier des meilleurs cours existants, soit comme une option à coûts réduits pour la formation continue des professionnels, en utilisant des outils de télécommunications. [S2]

□ **Formation :**  
Installation de salle multimédia pour la formation aux langues avec connexion à des sites étrangers. [S2]

□ **Santé et télémédecine :**  
Téléconsultation : consultation, diagnostic et suivi du patient à distance

Télé-Expertise : demande d'un deuxième avis à un médecin référent

Téléassistance à domicile

Téléassistance des voyageurs isolés : nautisme, montagne, trekking

Télé-médico-social : encadrement du patient maintenu à domicile

Télétransmission : transferts d'informations médicales

Télé-Radiologie : interprétation d'examens radiologiques à distance

Télé-Chirurgie : opération chirurgicale assistée à distance par ordinateur

Télé-Staff : réunion de professionnels de santé en visioconférence. [S2]

## 2.3 Caractéristiques de la vidéo :

Les principales caractéristiques de la vidéo sont : son nombre d'images par seconde, son entrelacement, sa résolution d'affichage, son allongement, la couleur espace et de bits par pixel, sa qualité et son début.

### 2.3.1. Nombre d'images par seconde

Le nombre d'images fixes par unité de temps de la vidéo, varie de six à huit images par seconde (*images / s*) pour les anciens appareils mécaniques à 120 ou plus d'images par seconde pour les nouveaux appareils photo professionnels. PAL (Europe, Asie, l'Australie, etc) et SECAM (France, Russie, certaines régions d'Afrique, etc) précisent les normes 25 images / s, tout en NTSC (USA, Canada, Japon, etc) spécifie 29,97 trame / s. Le film est tourné au taux plus lent cadre de 24photograms / s, ce qui complique un peu le processus de transfert d'un film cinématographique à la vidéo. Le taux de trame minimum pour atteindre l'illusion d'une image en mouvement est d'environ quinze cadres par seconde.

### 2.3.2. Entrelacement :

La vidéo peut être entrelacé ou progressif . L'entrelacement a été inventé comme un moyen d'atteindre une bonne qualité visuelle dans les limites d'une largeur de bande étroite. Les lignes de balayage horizontal de chaque trame entrelacée sont numérotées consécutivement et divisé en deux champs: le champ impair (champ supérieur) constitué par les lignes impaires-numérotés et le champ même (champ inférieur), composé des lignes numérotées-même. NTSC, PAL et SECAM sont entrelacés formats. vidéo spécifications résolution abrégées comportent souvent un **i** pour indiquer l'entrelacement. Par exemple, le format vidéo PAL est souvent spécifié comme **576i50**, où **576** indique la résolution verticale, **i** indique l'entrelacement, et **50** indique 50 champs (demi-images) par seconde..

Dans les systèmes à balayage progressif, chaque période de rafraîchissement des mises à jour toutes les lignes de balayage. Le résultat est une meilleure résolution spatiale et un manque de divers artefacts qui peuvent faire parties d'une image fixe semblent se déplacer ou clignotant.

Une procédure connue sous le nom de désentrelacement peut être utilisé pour convertir un flux entrelacé, comme analogique, DVD, ou par satellite, d'être traitées par des dispositifs de balayage progressif, tels que les téléviseurs LCD , les projecteurs et les écrans à plasma.

Désentrelacement ne peut pas produire une qualité vidéo équivalente à vrai source de matière balayage progressif.

### 2.3.3. Résolution de l'affichage :

La taille d'une image vidéo est mesurée en pixels pour la vidéo numérique, ou lignes de balayage horizontal et vertical des lignes de la résolution pour la vidéo analogique. Dans le domaine numérique (par exemple DVD)-télévision à définition standard ( SDTV ) est spécifié comme **720/704/640 × 480i60** pour le NTSC et **768/720 × 576i50** pour le PAL ou SECAM résolution. Toutefois, dans le domaine analogique, le nombre de lignes de balayage visible reste constante (486 NTSC/576 PAL), tandis que la mesure horizontale varie en fonction de la qualité du signal: environ 320 pixels par ligne de balayage pour la qualité de magnéscope, 400 pixels pour les émissions de télévision, et 720 pixels pour les sources DVD. Les proportions sont conservées en raison du non-carré "pixels".

Le nouveau téléviseur haute définition ( TVHD ) sa résolution atteint jusqu'à **1920 × 1080p60**, soit 1920 pixels par ligne de balayage par 1080 lignes de balayage, progressive, à 60 images par seconde.

La résolution vidéo pour la 3D-vidéo est mesurée en voxels (élément d'image de volume, ce qui représente une valeur dans un espace tridimensionnel). Par exemple la résolution **512 × 512 × 512** voxels, qui sert maintenant de simples 3D-vidéo, peuvent être affichées même sur certains PDA .

### 2.3.4 .Allongement :

Il décrit les dimensions des écrans vidéo et d'éléments d'image vidéo. Tous les formats vidéo les plus populaires sont rectilignes, et peut donc être décrite par un rapport entre la largeur et la hauteur. Le format d'un écran de télévision traditionnelle est de 4:3, soit environ 1.33:1. Les téléviseurs haute définition utilisent un ratio de 16:9, soit environ 1.78:1. Le rapport d'aspect d'un plein de 35 mm du film avec la bande son (aussi connu comme le ratio de l'Académie ) est 1.375:1.

Ratios où la hauteur est plus grande que la largeur sont rares dans une utilisation quotidienne, mais peuvent avoir une application dans les systèmes informatiques où l'écran peut être mieux adapté à une disposition verticale. Le grand rapport d'aspect de 3:4 plus courantes est appelée mode portrait et il est créé par rotation physiquement le dispositif d'affichage 90 degrés de



la position normale. D'autres rapports d'aspect de grande taille comme 9:16 sont techniquement possible mais rarement utilisé.

Pixels sur les écrans d'ordinateur sont généralement carré, mais pixels utilisés dans la vidéo numérique ont souvent des non-carrés aspect ratios, tels que ceux utilisés dans les variantes de PAL et NTSC de la norme CCIR 601 de vidéo numérique, et l'écran large anamorphique formats correspondants. Par conséquent, une image NTSC DV, ce qui est de 720 pixels par 480 pixels est affichée avec le ratio d'aspect de 4:3 (qui est la norme de télévision traditionnelle) si les pixels sont minces et s'affiche avec le ratio d'aspect de 16:9 (qui est le format écran large anamorphique) si les pixels sont gros.

### 2.3.5. Couleur espace et de bits par pixel :

Le nom du modèle de couleur décrit la représentation des couleurs vidéo. YIQ a été utilisé à la télévision NTSC. Il correspond à la YUV schéma utilisé dans la télévision NTSC et PAL et le YDbDr schéma utilisé par la télévision SECAM.

Le nombre de couleurs distinctes pouvant être représenté par un pixel dépend du nombre de bits par pixel (bpp). Une façon courante pour réduire le nombre de bits par pixel en vidéo numérique par sous-échantillonnage de chrominance (par exemple, 04:04:04 , 04:02:02 , 04:02:00 / 04:01:01 ).

### 2.3.6. Qualité vidéo :

La qualité vidéo peut être mesurer avec des paramètres formels tels que le PSNR ou avec une qualité vidéo subjective en utilisant l'observation d'experts.

La qualité vidéo subjective d'un système de traitement vidéo peuvent être évaluer comme suit:

- Choisissez les séquences vidéo (la *SRC*) à utiliser pour les tests.
- Choisissez les paramètres du système pour évaluer (*CDH*).
- Choisissez une méthode de test pour présenter des séquences vidéo à des experts et de recueillir leurs votes.
- Inviter un nombre suffisant d'experts, de préférence pas moins de 15.
- Réaliser des tests.
- Calculer la moyenne des notes pour chaque *HRC* sur la base d'experts évaluations.

Beaucoup méthodes subjective de la qualité vidéo sont décrites dans l'UIT-T BT.500 recommandation. Une des méthodes normalisées est Double Stimulus Impairment Scale (DSIS). En DSIS, chaque expert vue une vidéo de référence *irréprochable* suivie par une version *altérée* de la même vidéo. L'expert taux alors la vidéo *avec facultés affaiblies* en utilisant une échelle allant de «déficiences sont imperceptibles» à «les déficiences sont très ennuyeux».

### 2.3.7. Le débit :

Le débit est une mesure du taux de la teneur des informations dans un flux vidéo. Il est quantifié en utilisant l'unité **bits par seconde** (bit / s ou bps) ou mégabits par seconde (Mbit / s). Un taux plus élevé permet une meilleure qualité vidéo . Par exemple VideoCD, avec un débit d'environ 1 Mbit / s, est de moins bonne qualité que le DVD, avec un débit d'environ 5 Mbit / s. HD (High Definition Digital Vidéo et TV) a une qualité encore plus élevée, avec un débit d'environ 20 Mbit / s.

Un débit binaire variable (VBR) est une stratégie visant à maximiser la qualité vidéo visuel et minimiser le taux de bits. Sur les scènes de mouvement rapide, un débit binaire variable utilise plus de bits que ce n'est sur des scènes au ralenti d'une durée semblable encore atteint une qualité visuelle cohérente. Pour un temps réel et non-tamponné répartition d'une vidéo lorsque la bande passante disponible est fixée, par exemple un débit binaire constant (CBR) dans la visioconférence rendu sur les canaux d'une bande passante fixée doit être utilisé.



La résolution, terme souvent confondu avec la "définition", détermine par contre le nombre de points par unité de surface, exprimé en *points par pouce* (**PPP**, en anglais **DPI** pour *Dots Per Inch*); un pouce représentant 2.54 cm. La résolution permet ainsi d'établir le rapport entre le nombre de pixels d'une image et la taille réelle de sa représentation sur un support physique. Une résolution de 300 dpi signifie donc 300 colonnes et 300 rangées de pixels sur un pouce carré ce qui donne donc 90000 pixels sur un pouce carré. La résolution de référence de 72 dpi nous donne un pixel de 1"/72 (un pouce divisé par 72) soit 0.353mm, correspondant à un *point pica* (unité typographique anglo saxonne). [S5]

### 3.2.4 Le codage de la couleur :

Pour représenter informatiquement une image, il suffit de créer un tableau de pixels dont chaque case contient une valeur. La valeur stockée dans une case est codée sur un certain nombre de bits déterminant la couleur ou l'intensité du pixel, on l'appelle profondeur de codage (parfois *profondeur de couleur*). Il existe plusieurs standards de codage de la profondeur : bitmap noir et blanc, bitmap 16 couleurs ou 16 niveaux de gris, bitmap 256 couleurs ou 256 niveaux de gris, palette de couleurs (colormap) et Couleurs vraies (true color). [S5]

### 3.2.5 Luminance :

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet. [KS, 96]

### 3.2.6 Contraste :

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image.

Le contraste est défini en fonction des luminances de deux zones d'image.

Si L1 et L2 sont les degrés de luminosité respectivement de deux zones voisines A1 et A2 d'une image, le contraste C est défini par le rapport : 
$$C = \frac{L1-L2}{L1+L2} . [KS, 96]$$

### 3.2.7 Format :

Un format d'image est une représentation informatique de l'image, associée à des informations sur la façon dont l'image est codée et fournissant éventuellement des indications sur la manière de la décoder et de la manipuler. [S7]

Il existe un grand nombre de formats. Voici les plus utilisés :

Format	Compression des données	Nb de couleurs	Affichage progressif	Usage
<b>BMP</b>	Non compressé	de 2 à 16 millions	non	Image non dégradée mais très lourde. Stokage
<b>JPEG</b>	Réglable, avec perte de qualité. Plus la compression est importante, plus l'image est dégradée.	16 millions	oui	Tous usages, selon compression. Images "naturelles".
<b>GIF</b>	Oui, sans perte de qualité	de 2 à 256 avec palette.	oui	Logos et Internet.Supporte les animations et la transparence
<b>TIFF</b>	Réglable, au choix sans perte ou avec perte de qualité	16 millions	non	Tous sauf Internet
<b>PNG</b>	Oui, sans perte de qualité	de 2 à 256 ou 16 millions	oui	Tous, recommandé Internet mais incompatible avec les navigateurs anciens. Supporte la transparence.

Tableau 1.1 : Les principaux formats d'une image[S6]

## 4. Séparation avant plan \arrière plan :

### 4.1 Avant plan :

- la zone située entre l'appareil-photo et le sujet principal.
- la partie de la scène qui est près de ou devant de l'appareil-photo.
- la position ou les objets importants et remarquables dans l'image. [NT, 09]

### 4.2 Arrière-plan :

- la partie de la scène apparaissant derrière le sujet principal de la photo.
- le plan effacé et insignifiant en arrière d'un premier plan ou situé au fond dans une perspective. [NT, 09]

#### 4.3 Séparation avant-plan/arrière-plan par couleur :

- **Pré-conditions**

- \_ On doit assurer que les objets les plus importants sont plus saillants que l'arrière-plan.
- \_ On doit mettre le sujet près du centre de l'image et/ou font le sujet plus grand que d'autres objets dans l'image et/ou mettent le sujet dans le foyer tranchant.

- **Principe :**

Cette approche utilise la couleur des images pour proposer une méthode de segmenter automatiquement l'avant-plan et l'arrière-plan. Elle concerne l'élimination de l'arrière-plan dans l'image.

- **Avantage :**

- \_ On a choisi la couleur aux marges de l'image pour détecter l'avant-plan parce que la couleur est un facteur très important et significatif de caractéristiques visuelles de l'image.
- \_ Méthode est rapide et simple parce que le traitement de couleur demande peu de temps et de calculs.
- \_ S'accorde avec la plupart d'images spécifiques.

- **Inconvénient :**

- \_ Ne convient qu'à la base d'images particulières parce qu'elle ne rend pas de précision dans beaucoup d'images normales.
- \_ Quelques parties des objets ou tout l'objet de l'avant-plan peuvent être éliminées parce qu'elles ont le même couleur de l'arrière-plan. [NT, 09]

#### 4.4 Séparation avant-plan/arrière-plan par contours :

- **Pré-conditions**

- \_ Arrière-plan relativement lisse et uniforme dont les couleurs sont petit à petit variées ou dont la texture est mince.
- \_ On place un petit nombre d'objets de l'avant-plan au milieu de l'image dont le contraste est suffisant à l'égard de l'arrière-plan.

- **Principe :**

C'est une méthode adaptative se basant sur les contours pour localiser le profil (outline) des régions de l'avant-plan.

- **Avantage :**

0MFD 0MFD

\_ On peut détecter des régions de l'image en se basant sur les contours et séparer l'avant-plan et l'arrière-plan en même temps.

\_ Utilise la segmentation par région avec le bruit de l'avant plan.

- ***Inconvénient :***

\_ L'intégration de deux résultats de segmentation peut causer quelques problèmes.

\_ Convient seulement au type d'images avec l'arrière-plan assez lisse et uniforme et avec les objets de l'avant-plan se trouvent au milieu de l'image dont le contraste est suffisant à l'égard de l'arrière-plan. [NT, 09]

0MFD 0MFD

#### **4.5 Séparation avant-plan/arrière-plan par position :**

- ***Pré-conditions***

\_ S'intéresse seulement aux images dont des objets de l'avant-plan sont visuellement évidents et distincts de l'arrière-plan et des objets voisins (avoir des propriétés visuelles comme la couleur, la texture qui sont différents de l'arrière-plan et des objets voisins).

\_ Les objets doivent avoir la taille significative en comparaison de l'image.

- ***Principe :***

Intégration de la segmentation de l'image et la séparation entre l'avant-plan et l'arrière-plan dans un système unitaire, elle se concentre à utiliser la couleur pour segmenter l'image en régions ensuite utilise les caractéristiques visuelles pour regrouper les régions en une « grappe », enfin, les relations entre des grappes de l'avant-plan et de l'arrière-plan sont inférées.

- ***Avantage :***

\_ Applicable dans la reconnaissance des objets.

- ***Inconvénient :***

\_ Ne soit pas juste et suffisante dans tous les cas. [NT, 09]

#### 4.6 Séparation avant-plan/arrière-plan par ondelettes :

- *Pré-conditions :*

- \_ Il y a seulement des objets de l'avant-plan dans le foyer tranchant dans les images .
- \_ Les objets de l'arrière-plan sont flous.

- *Principe :*

Cette approche sépare les images de profondeur de champ faible, on détecte des objets de l'avant-plan en se basant sur la fréquence des ondelettes à multi résolution et sur les méthodes statistiques.

- *Avantage :*

- \_ Fonctionne très bien car elle s'intéresse à l'essence de l'image (l'ensemble de signaux).

- *Inconvénient :*

- \_ Elle est délimitée dans une base d'images particulière. Ce sont des images aux faibles profondeurs de champ ou des images avec une partie nette et une autre partie floue. [NT, 09]

#### 5. Les méthodes de détection et de classification :

Dans l'objectif d'analyser une image, la segmentation est alors une étape primordiale. On regroupe de façon usuel les méthodes de segmentation en quatre groupes.

##### 5.1 Approche globale de l'image (pixellaire) :

Le principe consiste à regrouper les pixels selon leurs attributs (intensité, contraste, vecteur moment) sans tenir compte de leur localisation au sein de l'image. Cela permet de construire des classes de pixels ; les pixels adjacents, appartenant à une même classe, forment alors des régions. L'objectif consiste à affecter à chaque pixel de l'image une étiquette  $e_i$  indiquant la classe (région) la plus proche (mesure de distance entre le pixel et le centre de gravité de la classe). [AP, 07]

Parmi les approches pixellaire, on distingue plusieurs méthodes :

##### 5.1.1 Les méthodes à seuils

l'une des méthodes de segmentation les plus simples à implémenter est Le seuillage d'histogramme ( thresholding) basé sur l'hypothèse forte qu'à chaque mode de l'histogramme correspond une région.

-Séparer les éléments d'une image en se basant uniquement sur l'intensité de niveaux de gris des éléments la constituant.



- Repérer les modes de l'histogramme (zones de forte densité) et rechercher les frontières entre ces modes dans les zones de faible densité.
- Utilisable juste dans le cas d'images peu bruitées, constituées d'objets d'intensités homogènes et significativement différentes d'un objet à l'autre.
- On obtient une séparation de l'histogramme en deux milieux uniquement si on utilise un seuillage simple. [AP, 07]
- Utiliser seuillage multiple pour les cas plus complexes d'histogramme.
- Recourir à un seuillage non uniforme pour traiter des images représentant des objets d'intensités homogènes uniquement localement.

### 5.1.2 La classification (clustering)

Cette méthode consiste à :

- regrouper et à classer les pixels d'une image en classes en fonction de leurs propriétés.
- associer un vecteur d'attributs à chaque point de l'image.
- Effectuer la classification sur ces vecteurs d'attributs de façon à aboutir à un nombre restreint de régions homogènes au sein de l'image.

Il existe deux grandes tendances : la classification supervisée, basée sur l'apprentissage de propriétés discriminantes sur un échantillon de données déjà classées, et la classification non supervisée basée sur une mesure de distance entre les vecteurs d'attributs. Nous présentons ici trois méthodes de classification qui nous paraissent être les plus utilisées :

- *Les k-plus-proches-voisins*

Cette approche supervisée, nécessitant une phase d'apprentissage, va classer chaque pixel de l'image en fonction de la classe d'appartenance des pixels de son voisinage (des k voisins plus proches en terme de distance entre leurs attributs et /ou leurs localisations spatiales). Contrairement à l'algorithme des k-moyennes, cette méthode est robuste même si les régions distinctes s'interpénètrent. Cette approche ne tenant toutefois pas compte des interactions spatiales entre pixels d'une même image, Cocquerez et Philipp ont proposé une méthode de relaxation permettant de corriger la présence de pixels isolés. [AP, 07]

- *Les k-moyennes*

Méthode non supervisée, elle nécessite cependant de connaître, *a priori*, le nombre de classes de l'ensemble d'arrivée. Les étapes à suivre sont :

- Choisir « n » centres de gravité de façon aléatoire.
- Affecter à chaque pixel l'étiquette de la classe la plus proche.
- Recalculer les centres de gravité.

-Répéter le processus jusqu'à stabilisation des centres de gravité.

Comme inconvénients de cette méthode on a : la sensibilité à l'initialisation des centres de gravité (risque de solution locale pas forcément optimale), en plus elle n'est pas efficace pour l'analyse d'images présentant des régions de formes quelconques. [AP, 07]

- *Réseaux de neurones*

Méthode supervisée nécessitant une phase contraignante d'apprentissage, les réseaux de neurones sont capables d'établir des relations entre les données en entrée du réseau et le résultat en sortie.

-Elle peut retourner la classe à laquelle appartient le pixel à partir de son niveau de gris et d'une connaissance du voisinage.

-Système difficile à mettre en place.

-Risque de sous et sur échantillonnage. [AP, 07]

## 5.2 Approche région :

L'objectif de ces méthodes consiste toujours à partitionner l'image traitée en régions homogènes ; chaque objet de l'image pouvant être ainsi constitué d'un ensemble de régions.

-On doit ajouter un critère de proximité géographique au critère d'homogénéité pour produire des régions volumineuses et éviter une division parcellaire des régions.

-Chaque pixel de l'image reçoit une étiquette lui indiquant son appartenance.

-les méthodes de cette approche restent sensibles à beaucoup de paramètres comme le sens de parcours des diverses régions ou encore l'initialisation des germes

On distingue deux familles d'algorithmes pour l'approche région : les méthodes de croissance de régions qui agrègent les pixels voisins selon le critère d'homogénéité; les méthodes qui fusionnent ou divisent les régions en fonction du critère choisi. [AP, 07]

### 5.2.1 La croissance de région :

La croissance s'effectue à partir de pixels initiaux appelés « germes » qui peuvent être sélectionnés de façon aléatoire ou automatique.

-les pixels adjacents à la région sont étudiés lors d'une itération du processus.

-le pixel sera agrégé dans la région s'il vérifie les conditions d'homogénéité.

-Les pixels non intégrés aux régions peuvent générer eux-mêmes de nouvelles régions ou être quand même assimilés à la région la plus proche.

-méthode très sensible à la variation d'un de ces paramètres et le résultat final en est fortement dépendant.

-bien adaptée aux séquences d'images présentant de faibles variations comme par exemple les images d'une acquisition vidéo. [AP, 07]

### 5.2.2 Les méthodes de décomposition et fusion (« *Splitting and Merging* »)

Le principe consiste en une alternance de phases de division et de fusion de régions jusqu'à optimiser un critère d'homogénéité choisi préalablement. Pour l'étape de fusion, on recherche dans les couples de régions possibles (régions adjacentes quasi similaires), ceux candidats à un possible regroupement. Chaque couple retenu est noté en fonction de l'impact qu'aurait sa fusion sur le critère d'homogénéité global. Les couples les mieux notés sont alors fusionnés. L'étape de décomposition (division) agit de façon opposée ; les régions les moins homogènes sont alors divisées en régions plus petites. Le processus de division peut, dans le cas extrême, aboutir à un maillage de l'image sous forme de pixels indépendants. Le résultat final est obtenu lorsque la condition d'arrêt prédéfinie est atteinte (nombre de régions, contraste) ou lorsque les notes attribuées aux couples candidats à la fusion n'évoluent plus significativement. Dans l'image est d'abord partitionnée en utilisant les lignes de partage des eaux avant la fusion des régions similaires ainsi obtenues. [AP, 07]

Deux des méthodes de décomposition et fusion de régions les plus utilisées reposent sur le diagramme de Voronoï et l'utilisation d'un arbre quaternaire.

- *Diagramme de Voronoï*

L'utilisation du diagramme de Voronoï permet un partitionnement de l'image en polygones. A l'initialisation, le diagramme est construit sur un ensemble de germes sélectionnés de façon aléatoire sur l'image via un processus de Poisson. L'application des phases de fusion et de décomposition permet la suppression de régions (et de germes) superflus ainsi qu'une décomposition des régions hétérogènes au sens du critère adopté initialement. [AP, 07]

- *Arbre quaternaire (« quadtree »)*

Le principe consiste à diviser l'image initiale en régions rectangulaires répondant toutes au critère d'homogénéité avant d'appliquer la phase de fusion. Les étapes de cette méthode sont comme suite :

- Prendre l'image comme région initiale.
- Subdiviser en quatre la région si elle ne répond pas au critère d'homogénéité pour construire l'arbre.
- Répéter l'opération sur les quatre régions filles si besoin.
- Regrouper ensemble les feuilles de l'arbre ayant des caractéristiques similaires durant la phase de fusion. [AP, 07]

### 5.3 Approche contour :

Les approches contour cherchent à trouver les frontières des régions, on se base sur des mesures de gradients au sein de l'image, elles consistent en une étude locale de recherche de discontinuités, Les contours extraits via les calculs des dérivées première et seconde ne sont généralement pas fermés et /ou continus alors il est nécessaire d'associer une méthode de suivi et /ou de fermeture des contours. [AP, 07]

On discerne trois grandes familles d'approches :

#### 5.3.1 Les méthodes de calculs de gradients discrets

L'utilisation des filtres tels que ceux de Sobel , Prewitt ou encore Roberts permet de calculer les gradients locaux en tout point de l' image. Ils permettent de discrétiser le calcul des dérivées. Ces méthodes, bien que rapides à implémenter, ne sont malheureusement pas robustes au bruit. [AP, 07]

#### 5.3.2 Les détecteurs de contour

Après le calcul de la norme et la direction du gradient en chaque pixel de l'image ces méthodes extraient des contours d'un seul pixel d'épaisseur en sélectionnant les maxima locaux des normes des gradients. On peut citer les détecteurs de Canny, Deriche, Marr -Hildreth ou celui plus récent de Bao, Pour les opérateurs de Canny et de Deriche, une variable supplémentaire permet d'ajuster la sensibilité de la détection au détriment de la précision spatiale du contour et *vice versa*. On peut de plus noter que la détection de contours dans les zones à forts gradients ne correspond pas forcément à un objet. Dans ce cas, un post -traitement peut s'avérer efficace. [AP, 07]

Ces méthodes, de mi se en œuvre simple, donnent de bons résultats pour l'analyse d'image représentant des objets non texturés, fortement contrastés, aux contours marqués.

#### 5.3.3 Les modèles déformables

Ils sont introduits par Kass et Terzopolous et aussi connus sous les noms de « snakes » ou « contour s actifs ». Un contour actif agit comme un élastique, il modifie sa forme de façon à venir épouser celle de la frontière recherchée. Ces méthodes sont peu sensibles au bruit et gèrent de façon efficace les contours irréguliers. Ils appartenant au domaine continu, les contours obtenus ont une précision supérieure à la résolution de l'image étudiée.

Ces méthodes sont beaucoup employées dans l'imagerie médicale. Leur inconvénient majeur réside dans l'initialisation du contour actif et le choix des poids alors ils nécessitent d'avoir une connaissance a priori, de l'environnement étudié et de l'objet recherché. [AP, 07]

#### 5.4 Approche coopérative

On trouve les approches hybrides ou mixtes, ces approches représentent une combinaison des résultats de diverses méthodes afin d'en ajouter leurs avantages. Alors on prend la localisation des frontières, des contours, non continus par une approche contour et on la joignant une approche région dont les caractéristiques sont l'obtention de zones fermées et homogènes, encore on peut ainsi pallier les faiblesses de chacune des techniques (la faible précision du contour dans l'approche région et l'obtention de régions non fermées dans l'approche contour). . [AP, 07]

On distingue différentes catégories de méthodes mixtes :

##### 5.4.1 Par fusion de résultats obtenus avec diverses méthodes

Citons l'exemple d'une coopération entre une méthode basée croissance de régions et une méthode par modèle déformable. La compétition entre régions fait se développer une région de couleur homogène tout en limitant les contours irréguliers.

##### 5.4.2 Par utilisations successives des approches contour et région

La méthode région va alors utiliser la carte de contours obtenue via l'application d'une approche contour. Elle permet d'accroître la précision des frontières entre régions. Un exemple de cet te approche peut être l'utilisation d'une approche par croissance de régions en forçant les frontières à épouser la car te des contours obtenue par une des approches contour. . [AP, 07]

##### 5.4.3 Par utilisation simultanée de méthodes basées contour et région

Cette méthode propose une amélioration de la méthode de division-fusion liée à l'arbre quaternaire en y ajoutant une méthode similaire aux modèles déformables. Les deux premières phases de l'algorithme sont en effet celles de la division- fusion : création de l'arbre quaternaire sur un critère d'homogénéité puis fusion des régions voisines selon ce même critère. La troisième phase correspond à l'utilisation d'un modèle déformable tenant compte de la norme du gradient local et de la courbure du contour.

www.pdf-converter-software.com

Cette méthode propose l'avantage de lisser les contours obtenus par la simple application de la méthode de division-fusion du « quadtree » mais reste cependant fortement liée à l'arbre quaternaire et nécessite le réglage de nombreux paramètres. [AP, 07]

## 6. Les méthodes d'extraction fond/forme

Nous présenterons ici trois méthodes pour isoler les objets en mouvement, souvent appelées méthodes d'extraction fond/forme. Ces trois méthodes sont : Mixture de gaussiennes, Codebook 2 layers et Vumètre.

### 6.1 Mixture de gaussiennes :

Dans cette approche, chaque pixel est modélisé par une mixture de  $N$  gaussiennes,  $2 \leq N \leq 5$ . Pour  $n = 1, \dots, N$ , un élément de la mixture de gaussiennes est représenté par une moyenne  $\mu_n$ , un écart type  $\sigma_n$ , et un Poids  $\alpha_n$  ( $\sum_n \alpha_n = 1$ ). On peut remarquer que  $\alpha_n$  est réduit à un scalaire.

Pour une nouvelle image traitée, la mixture de gaussiennes (pour tous les pixels) est mise à jour pour expliquer correctement les couleurs affichées par chaque pixel. Pour faire ceci, à un instant  $t$ , on considère que le modèle  $M_t$  généré pour chaque pixel à partir des mesures  $\{Z_0; Z_1, \dots, Z_{t-1}\}$  est correct. La vraisemblance pour qu'un pixel appartienne au fond est :

$$P(Z_t | M_t) = \sum_{n=1}^{n=N} \alpha_n N(\mu_n, \Sigma_n)$$

$$N(\mu_n, \Sigma_n) = \frac{1}{2\pi^{d/2} |\Sigma_n|^{1/2}} e^{-1/2(Z_t - \mu_n)^T \Sigma_n^{-1} (Z_t - \mu_n)}$$

Avec  $d$  la dimension de l'espace de couleurs de la mesure  $Z_t$ .

Pour mettre à jour le modèle, on associe d'abord la mesure  $Z_t$  à une gaussienne  $n$  si :

$$\|Z_t - \mu_n\| < K \sigma_n$$

Où  $K$  vaut 2 ou 3. L'opérateur  $<$  est vrai si toutes les composantes du vecteur à gauche sont inférieures à  $K \sigma_n$ .

Cette mesure représente le fond si la gaussienne  $n$  explique le fond de la scène. En fait, le poids  $\alpha_n$  est élevé.

Cette gaussienne est alors mise à jour :

$$\hat{\alpha}_n \leftarrow (1 - \delta)\hat{\alpha}_n + \delta$$

$$\hat{\mu}_n \leftarrow (1 - \delta)\hat{\mu}_n + \delta Z_t$$

$$\hat{\sigma}_n^2 \leftarrow (1 - \delta)\hat{\sigma}_n^2 + (Z_t - \hat{\mu}_n)^T(Z_t - \hat{\mu}_n)$$

Avec  $\delta$  le coefficient d'apprentissage. Il représente la vitesse d'adaptation du modèle. Pour toutes les autres gaussiennes  $n \neq \hat{n}$ , la moyenne et la variance ne sont pas modifiées, mais :

$$\alpha_n \leftarrow (1 - \delta)\alpha_n$$

Si le test 3 échoue, le pixel est associé au 1er plan. La gaussienne ayant le plus petit poids est réinitialisée avec la mesure actuelle :

$$\alpha_n = \delta$$

$$\mu_n = Z_t$$

$$\sigma_n^2 = \bar{\sigma}^2$$

Avec  $\bar{\sigma}^2$  une variance élevée. Ces affectations sont aussi appliquées pour l'initialisation de la mixture. [TGG, 99]

## 6.2 Codebook 2 Layers(CB2) :

Cette méthode, est très largement inspirée du codebook. Mais elle en diffère, en utilisant deux codebooks, bibliothèques de données pour chaque pixel contenant des informations pour modéliser le fond. Ceci a été réalisé de manière à pouvoir retenir des valeurs de pixels qui ont appartenu au fond, mais qui pourraient redevenir du fond, c'est typiquement le cas avec les mouvements de branches d'arbre.

Chaque codebook contient des éléments appelés codeword (CW) pour modéliser le fond de l'image, chacun des CW contient ces informations :

- $\mathbf{v}_i$  : valeur moyenne du pixel (R,V,B).
- $I_{max}$  : limite maximale d'intensité du CW.
- $I_{min}$  : limite minimale d'intensité du CW.
- $f$  : fréquence du CW (nombre d'occurrences).
- $\lambda$  : nombre maximal d'images où le CW ne correspond à aucun pixel.

- $p$  : première occurrence du CW.
- $q$  : dernière occurrence du CW.

Le principe est le meme qu'avec le codebook simple, mais avec deux codebooks par pixel : un principal appelé M, et un secondaire appelé H.

Le traitement se fait en 2 phases : une phase d'apprentissage qui sert à créer les codebooks principaux initiaux, et une phase de soustraction pour extraire le fond de la forme.

Pour chaque nouveau pixel  $x_t = (R, V, B)$ , son intensité  $I_t$  est calculée par  $I_t = \sqrt{R^2 + V^2 + B^2}$

La distorsion de couleur  $\delta$  entre ce pixel  $x_t = (R, V, B)$ , et un codeword  $c_i$  avec  $v_i = (\bar{R}_i, \bar{V}_i, \bar{B}_i)$  peut être calculé par :

$$\langle x_t, v_i \rangle^2 = (\bar{R}_i R + \bar{V}_i V + \bar{B}_i B)^2$$

$$\|v_i\|^2 = \bar{R}_i^2 + \bar{V}_i^2 + \bar{B}_i^2$$

$$\|x_t\|^2 = R^2 + V^2 + B^2$$

$$colordist(x_t, v_i) = \delta = \sqrt{\|x_t\|^2 - \frac{\langle x_t, v_i \rangle^2}{\|v_i\|^2}}$$

Un pixel  $x_t$  avec une intensité  $I_t$  correspond à un codeword  $c_i$  avec une valeur de pixel  $v_i$  et  $I_{min}, I_{max}$  si  $I_t$  est dans l'intervalle  $[I_{min}, I_{max}]$  et la distorsion de couleur  $\delta$  respecte  $\delta < \epsilon$ .

En phase d'apprentissage, seule la couche M est construite, H reste vide. Pour un nouveau pixel  $x_t$ , on cherche un CW dans M correspondant à  $x_t$ . Si on en trouve un, il est mis à jour avec  $x_t$ , sinon un nouveau CW est créé à partir de la valeur de  $x_t$ .

Un nouveau codeword est créé avec un pixel  $x_t$  de la façon suivante :

$$v_i \leftarrow (R, V, B)$$

$$I_{min} \leftarrow \max\{0, I_t - \alpha\}$$

$$I_{max} \leftarrow \min\{255, I_t + \alpha\}$$

$$f \leftarrow 1; \lambda \leftarrow t - 1; p \leftarrow t; q \leftarrow t$$

avec  $\alpha$  une valeur représentant une tolérance d'intensité.



Pendant la phase d'apprentissage, un codeword est mis à jour par un pixel  $\mathbf{x}_t$  comme ceci :

$$\bar{R} \leftarrow \frac{\bar{R} \times f + R}{f + 1} \text{ (de meme pour } V \text{ et } B)$$

$$I_{min} \leftarrow \frac{I - \alpha + f \times I_{min}}{f + 1}$$

$$I_{max} \leftarrow \frac{I + \alpha + f \times I_{max}}{f + 1}$$

$$f \leftarrow f + 1; \lambda \leftarrow \max\{\lambda, t - q\}; p \leftarrow p; q \leftarrow t$$

En phase de soustraction, pour un nouveau pixel  $\mathbf{x}_t$ , on cherche un CW dans M correspondant à  $\mathbf{x}_t$ . Si on en trouve un, il est mis à jour avec  $\mathbf{x}_t$  et le pixel est considéré comme appartenant au fond. Sinon, on cherche un CW correspondant à ce pixel dans H, si on en trouve un, on le met à jour avec  $\mathbf{x}_t$ , sinon, on en crée un nouveau dans H avec la valeur de  $\mathbf{x}_t$ .

Un CW est mis à jour comme précédemment dans la phase d'apprentissage à l'exception de  $I_{min}$  et  $I_{max}$  qui sont mis à jour de la façon suivante, avec  $\beta$  un coefficient pour changer la vitesse d'adaptation :

$$I_{min} \leftarrow (1 - \beta)(I_t - \alpha) + \beta \cdot I_{min}$$

$$I_{max} \leftarrow (1 - \beta)(I_t + \alpha) + \beta \cdot I_{max}$$

Ensuite, les modèles M et H sont affinés avec ces règles :

- Supprimer les CWs de H ayant  $\lambda > T_H$
- Déplacer les CWs restant plus de  $T_{add}$  dans H vers M
- Supprimer les CWs de M n'apparaissant pas plus longtemps que  $T_{delete}$ . [SF, 08]

### 6.3 Vumètre (VUM) :

Le vumètre par Goyat et al est un modèle non paramétrique, basé sur une estimation discrète de la probabilité de distribution. Il s'agit d'une approche probabiliste pour définir le modèle du fond.

Soit  $I_t$  une image à l'instant  $t$ ,  $\mathbf{y}_t(\mathbf{u})$  donne les valeurs RVB du pixel  $\mathbf{u}$ .

Un pixel peut prendre deux états, ( $\omega_1$ ) s'il appartient au fond, ( $\omega_2$ ) s'il appartient au 1er plan. Cette méthode essaye d'estimer  $p(\omega_1|y_t(u))$  Avec 3 composantes de couleur i (Rouge, Vert, Bleu), la fonction de densité de probabilité peut être approximée par :

$$p(\omega_1|y_t(u)) = \prod_{i=1}^3 p(\omega_1|y_t^i(u))$$

Avec :

$$p(\omega_1|y_t^i(u)) \cong K_i \sum_{j=1}^N \pi_t^{ij} \delta(b_t^i(u) - j)$$

où  $\delta$  est le symbole de Kronecker,  $b_t(u)$  donne le vecteur d'index de la classe associée à  $y_t(u)$ ,  $j$  est un index de classe, et  $K_i$  est une constante de normalisation permettant de garder à chaque instant :

$$\sum_{j=1}^N \pi_t^{ij} = 1$$

$\pi_t^{ij}$  est une fonction de masse discrète représentée par une classe.

A la première image ( $t = 0$ ), les valeurs des classes sont mises à  $\pi_0^{ij} = 1/N$  pour garder une somme à 1 comme dans l'équation 27. A chaque nouveau pixel, sa valeur correspond à une classe  $\pi_t^{ij}$ , son niveau est mis à jour de cette façon :

$$\pi_{t+1}^{ij} = \pi_t^{ij} + \alpha \cdot \delta(b_{t+1}^i(u) - j)$$

Après un certain nombre d'images, les classes modélisant le fond ont une valeur élevée. Pour pouvoir décider à quel moment un pixel appartient au fond ou non, un seuil T est défini.

Chaque nouveau pixel ayant une classe correspondante sous ce seuil sera détecté comme appartenant au fond.

En mode RVB, chaque pixel est modélisé par 3 vumètres (un par composante). Pour considérer un pixel comme fond, il doit être détecté comme fond par chaque vumètre.

Afin d'améliorer la détection et réduire des problèmes liés aux seuils entre deux classes, la valeur des classes au voisinage de la classe correspondante à un pixel est aussi mise à jour, mais de façon moindre.

Pour obtenir un bon apprentissage et une bonne adaptation de l'algorithme, il est nécessaire de bien choisir les paramètres (taux d'apprentissage  $\alpha$  et seuil  $T$ ). Ces valeurs peuvent être changées en fonction de la luminosité ou de la vitesse des véhicules suivis. [GCMT, 06]

## **7. Conclusion :**

Dans ce chapitre nous avons présenté un petit résumé sur la vidéo et l'image suivi par les différentes méthodes utilisées pour traiter les images d'une vidéo de surveillance. Premièrement nous avons parlé de la séparation avant plan \ arrière plan, ensuite nous avons présenté les méthodes de détection et de classification et nous avons terminé par les méthodes d'extraction fond \ forme.

Remove  
Please  
Registered  
With  
<http://www.pdf-converter-software.com>

## 1. Introduction

L'évolution de la société humaine a entraîné le développement parallèle de maux sociaux qui tendent à freiner son élan dans son parcours naturel vers la perfection. Seulement, il est inconcevable que ce parcours soit dénué d'obstacle : la criminalité organisée, le terrorisme, les vols, le banditisme, détournement de fonds...etc. Pour se défaire de ces obstacles il est devenu impératif de trouver des solutions.

Au début, la société a dû utiliser l'homme lui-même pour venir à bout de ces problèmes qui croissent plus en plus et demandent un nombre infini d'agent de surveillance. Alors, le besoin de solutions s'est imposé et la société a dû chercher des solutions que seule la technologie est en mesure de fournir.

De nos jours, les agents de sécurité ont montré leurs limites, car beaucoup de séquences visuelles peuvent leur échapper. La solution qui s'impose est la caméra de surveillance qui secondera l'agent dans plusieurs espaces vitaux. Cependant, comme l'être humain a des limites, sa création (la caméra) a aussi des limites.

Mais, n'empêche la vidéosurveillance joue un rôle important en permettant la surveillance, le suivi et la détection des individus dans des séquences vidéo. Toutefois, l'enregistrement des données reste une étape facile par rapport au traitement qui demande davantage de temps et nécessite d'être affiné.

## 2. Problématique

L'analyse du comportement humain est une tâche très importante dans diverses applications comme la vidéo surveillance, la recherche vidéo, les systèmes d'interaction homme machine, et les systèmes de diagnostics médicaux. L'un des problèmes les plus difficiles dans ce type d'analyse est la détection ensuite la segmentation des individus en mouvement.

- Le premier problème se situe dans la manière d'extraire les objets en mouvement par rapport à l'arrière-plan.
- Le second sera de pouvoir segmenter l'individu de tel sorte que celui-ci apparait dans son entier et ensuite pouvoir suivre les activités de ce dernier.

Et pour pouvoir régler ces problèmes nous avons opté pour la solution suivante schématisée par l'architecture du système conçue ainsi.

### 3. Architecture du système

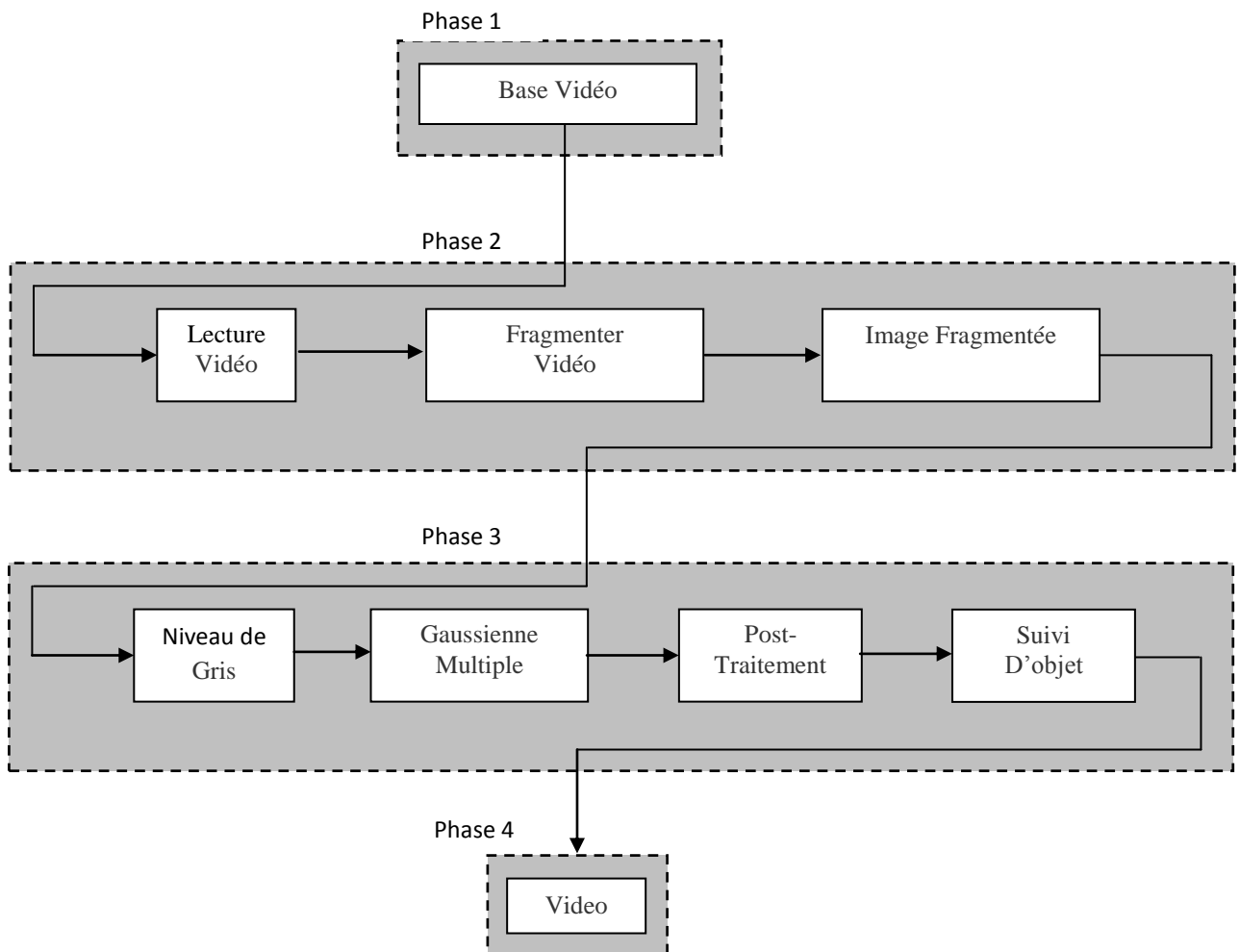


Figure 2.1 : Architecture du système



### 4.2.3 Image fragmentée

C'est une image fixe, résultat de la fragmentation de la séquence vidéo.

## 4.3 Phase 3

Cette phase se compose de quatre étapes essentielles :

### 4.3.1 Niveau de gris

Les images en couleurs, possédant trois composantes couleurs : Rouge Vert et Bleu, sont transformées en images possédants une seule composante grise pour gagner du temps et faciliter le travail. La valeur de cette dernière est calculée selon l'algorithme suivant :

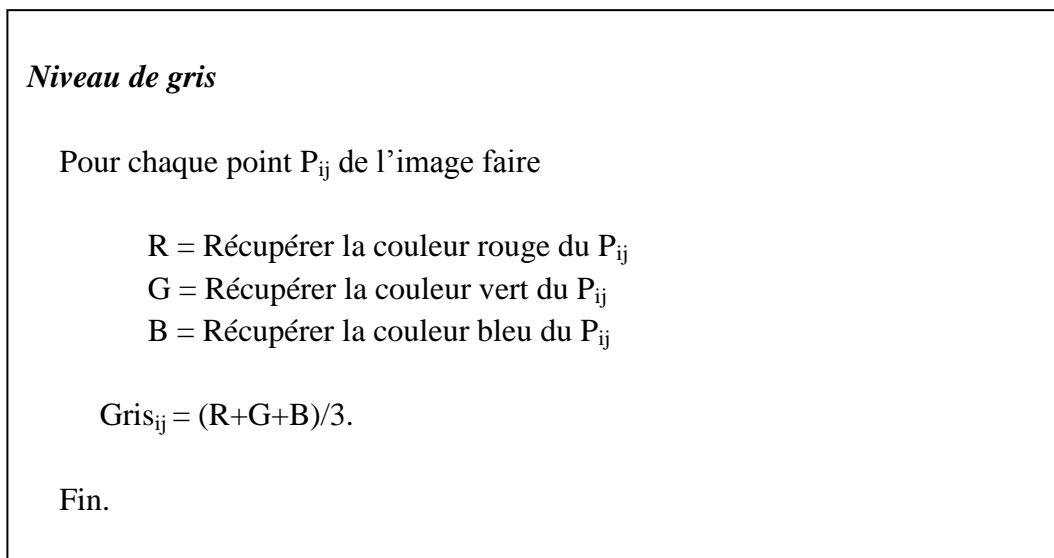


Figure 2.3 : Algorithme niveau de gris

### 4.3.2 Gaussienne multiple

Elle comporte deux étapes :

#### 4.3.2.a Apprentissage

Consiste à présenter à la machine une suite d'images comportant uniquement l'arrière plan dans le but que cette dernière puisse mémoriser cet arrière plan et pouvoir le reconnaître ultérieurement, et, ceci au moyen de l'algorithme des gaussiennes multiples.

#### 4.3.2.b Extraction de l'avant plan

Consiste à faire passer les images comportant l'individu objet de l'étude afin de séparer l'arrière plan de l'avant plan au moyen de l'algorithme des gaussiennes multiples.

L'opération se fait au moyen de l'algorithme qui suit dont voici l'explication des formules et indices utilisés:

➤ *Explication des indices*

$K$  : nombre des gaussiennes associées a chaque pixel.

$V_0$  : valeur initiale de départ.

$\alpha$  : taux d'apprentissage ou constante de temps.

$T$  : détermine la proportion de données minimum correspondant au fond.

$w_j$  : poids calculé pour chaque gaussiennes  $K$ .

$\mu_j$  : moyenne de la gaussienne  $K$ .

$\sigma_j$  : écart-type de la gaussiennes  $K$ .

$q_k$  : égal 1 pour la gaussienne qui correspond à l'arrière plan et 0 pour les autres.



www.pdf-converter-software.com

**Gaussienne multiple**

Variables de contrôles :  $K \quad V_0 \quad \alpha \quad T$

Initialisation :  $\forall j = 1..K \quad w_j=0 \quad \mu_j= Inf \quad \sigma_j= V_0 \quad c_j = 0$

Tant que nouveau pixel  $x(t)$

$$\forall_{j=1..K} p_j = \begin{cases} w_j g_j(x; \mu_j, \sigma_j) & \text{if } \frac{|x - \mu_j|}{\sigma_j} < T \\ 0 & \text{autres} \end{cases}$$

Si  $\sum_{j=1}^K p_j > 0$  Alors

Pour ( $k=1 ; k < K ; k++$ )

$$q_k = \frac{p_k}{\sum_j p_j}$$

Si Gagnant-Prendre-Tous Alors

$$q_k = \begin{cases} 1 & \text{if } k = \text{argmax}_j \{p_j\} \\ 0 & \text{Autres} \end{cases}$$

Fin Si

$$w_k(t) = (1 - \alpha) \cdot w_k(t - 1) + \alpha q_k$$

Si  $q_k > 0$  Alors

$$c_k = c_k + q_k \quad \varphi_k = q_k \cdot \left( \frac{1-\alpha}{c_k} + \alpha \right)$$

$$\begin{aligned} \mu_k(t) &= (1 - \varphi_k) \cdot \mu_k(t - 1) + \varphi_k \cdot x \\ \sigma_k^2(t) &= (1 - \varphi_k) \cdot \sigma_k^2(t - 1) + \varphi_k \cdot (x - \mu_k(t - 1))^2 \end{aligned}$$

Fin Si

Fin Pour

Sinon

$$\begin{aligned} \forall_{j=1..K} w_j(t) &= (1 - \alpha) \cdot w_j(t - 1) \\ k = \text{argmin}_j \{w_j\} \quad w_k &= \alpha \quad \mu_k = x \quad \sigma_k = V_0 \quad c_k = 1 \end{aligned}$$

Fin Si

Normalisation w

Fin tant que

Figure 2.4 : Algorithme des gaussiennes multiples.

➤ *Explication des formules*

$g_j(x; \mu_j, \sigma_j)$  : Densité de probabilité d'une gaussienne.

$\frac{|x - \mu_j|}{\sigma_j} < T$  : permet de calculer la distance entre le pixel  $x$  et chacune des gaussiennes.

$w_k(t) = (1 - \alpha) \cdot w_k(t - 1) + \alpha q_k$  : permet la mise à jour des poids des gaussiennes.

$\mu_k(t) = (1 - \varphi_k) \cdot \mu_k(t - 1) + \varphi_k \cdot x$  : permet la mise à jour des moyennes des gaussiennes.

$\sigma_k^2(t) = (1 - \varphi_k) \cdot \sigma_k^2(t - 1) + \varphi_k \cdot (x - \mu_k(t - 1))^2$  : permet la mise à jour des écarts-types des gaussiennes.

#### 4.3.2.c Post-traitement

Il comporte deux étapes :

##### 1- Dilatation automatique

Consiste à boucher les trous sur le corps de l'individu résultant de l'algorithme des gaussiennes multiples pour avoir une forme homogène sans perdre les caractéristiques morphologiques de l'objet, selon l'algorithme qui suit :

Masque1 Masque2 Masque3 Masque4 Masque5 Masque6 Masque7 Masque8 Masque9

Pour s'en passer de la dilatation ancienne qui a tendance a fausser les caractéristiques morphologiques de l'objet nous recourons a la dilatation automatique utilisant ces masques que nous avons testé et dont les résultats se montrent probants :

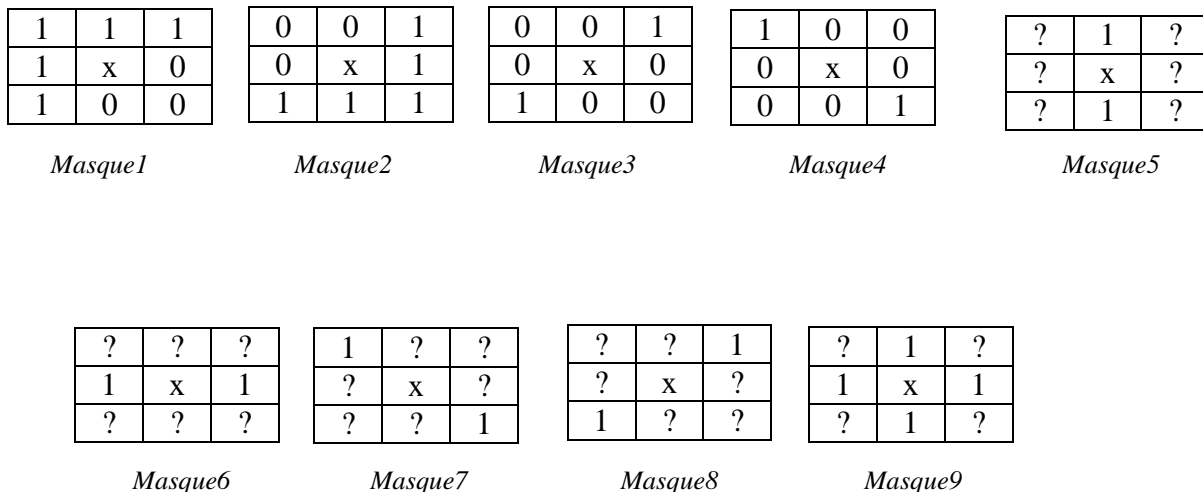


Figure 2.5 : Les masques utilisés dans la dilatation automatique.

1 : Signifie la couleur blanche.

0 : Signifie la couleur noire.

? : Signifie le pixel peut prendre la valeur 1 ou 0.

***Dilatation automatique***

Modification = vrai.

Tant que (Modification) faire

Modification = Faux.

Pour chaque point P<sub>ij</sub> de l'image faire

Récupérer les valeurs des voisinages de P<sub>ij</sub>.

Si (masque1 ou masque2 ou masque3 ou masque4 ou masque5 ou  
masque6) alors

Colorer P<sub>ij</sub> blanc.

Modification = vrai.

Fin.

Si le nombre des voisinages du P<sub>ij</sub> portant la valeur 1 = 3 alors

Si (masque7 ou masque8) alors

Colorer P<sub>ij</sub> blanc.

Modification = vrai.

Fin.

Fin.

Si le nombre des voisinages du P<sub>ij</sub> portant la valeur 1 = 4 alors

Si (masque7 ou masque8 ou masque9) alors

Colorer P<sub>ij</sub> blanc.

Modification = vrai.

Fin.

Fin.

Si le nombre des voisinages du P<sub>ij</sub> portant la valeur 1 ≥ 5 alors

Colorer P<sub>ij</sub> blanc.

Modification = vrai.

Fin.

Fin.

Fin.

Figure 2.6 : Algorithme de dilatation automatique.

## 2- Élimination des points isolés

Consiste à faire un balayage de l'image de façon à supprimer les points rebelles de façon automatique, selon l'algorithme qui suit :

### ***Éliminer les points isolés***

Modification = vrai.

Tant que (Modification) faire

    Modification = Faux.

    Pour chaque point  $P_{ij}$  de l'image faire

        Récupérer les valeurs des voisinages de  $P_{ij}$ .

        Si le nombre des voisinages du  $P_{ij}$  portant la valeur(0) = 8 alors

            Colorer  $P_{ij}$  noire.

            Modification = vrai.

        Fin.

    Fin.

Fin.

Figure 2.7 : Algorithme d'élimination des points isolés.

### 4.3.3 Suivi d'objet

C'est la localisation spatiale et temporelle d'un objet au cours d'une séquence vidéo, où un objet est un ensemble de pixels décrivant la même représentation sémantique (ex : une voiture, un visage, un personnage...). Schématiquement, le principe consiste à extraire de l'objet des caractéristiques que l'on va s'efforcer de retrouver à chaque image de la séquence vidéo. Selon l'algorithme qui suit :

www.pdf-converter-software.com

```
Histogramme  
  
Hist= vecteur de taille égale la largeur de l'image.  
  
Pour i= 1 à largeur faire  
    Nb=0 ;  
    Pour j= 1 à hauteur faire  
        Si le point Pij de l'image = 255 alors  
            Nb=Nb+1.  
        Fin.  
    Histi = Nb ;  
Fin.  
Fin.
```

Figure 2.8 : Algorithme de calcul de l'histogramme d'une image.

```
Suivi d'objet  
  
Calculer l'histogramme horizontal de l'image.  
Compter le nombre des objets horizontaux.  
Diviser l'image en sous images selon le nombre d'objets trouvés.  
  
Pour chaque sous image faire  
  
    Calculer l'histogramme vertical  
    Compter le nombre des objets verticaux.  
    Extraire les extrémités des objets verticaux.  
    Dessiner des rectangles qui englobent les objets.  
  
Fin Pour.
```

Figure 2.9: Algorithme traitant le suivi des objets dans une image.

#### 4.4 Phase 4

Cette phase permet la création d'une séquence vidéo à partir d'images résultants de l'algorithme de suivi d'objets.

##### **Créer vidéo**

Je dispose d'un fichier d'images.

Créer objet multimédia sur un chemin donné.

Tant que image existe faire

    Lire l'image à partir du fichier d'images.

    Harmoniser l'image avec l'objet multimédia déjà crée.

    Attribuer un temps  $t$  à chaque image.

    Mettre l'image dans un buffer media.

    Insérer l'image dans l'objet multimédia.

Fin.

Figure 2.10: Algorithme de création d'une vidéo.

## 5. Conclusion

Dans ce chapitre, nous nous sommes intéressés à la structure de notre application et nous avons donné les principaux algorithmes composant notre logiciel.

PDF CONVERTER



## 1. Introduction

Ce chapitre cherche à mettre en évidence les résultats obtenus suite aux tests réalisés sur vidéos différentes par les contextes et les objets. Les tests nous ont permis de mettre en évidence les différentes étapes réalisées par notre logiciel à savoir : la détection, la segmentation et le suivi d'un individu en mouvement dans son environnement vital. Notre objectif est d'arriver à optimiser notre application. En aucun cas nous ne pouvons négliger notre base vidéo que nous allons utiliser dans les passages qui suivront.

## 2. Base de vidéo

De notre base vidéo nous en avons choisis trois types de vidéos différentes par : le contexte, les fluctuations d'intensité lumineuse, la richesse en objets, le nombre et la nature des personnes, les mouvements provoqués par les éléments actifs de la nature tel que le vent, les nuages, les poussières, les bruits, les mouvements involontaires du photographe.

### 2.1 Vidéo de type milieu naturel ouvert

Cet environnement complexe est une multitude d'objets différents par la grosseur, la hauteur, la largeur, la couleur, l'architecture, l'enchevêtrement, la disposition et surtout de l'individu par rapport à l'ensemble. Voici une suite d'images vidéo, environnement complexe dont :

- a- Image extraite vidéo comportant uniquement l'arrière plan avant l'intégration de l'individu.
- b- Image issue du film vidéo avec arrière plan, individu et nouvel objet, en mouvement, détectable.
- c- Image issue du film vidéo avec arrière plan et plus d'un individu.





(a) Image de référence



(b) Image N° 494

Figure 3.2 : Images vidéos de type environnement simple.

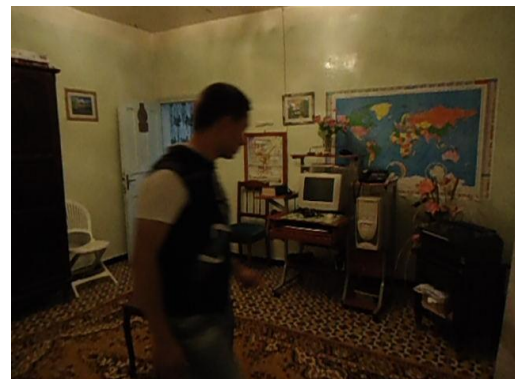
### 2.3 Vidéo dans un milieu type administratif

Avec objets multiples et caractéristiques dont bureau type micro ordinateur, dictionnaire, téléphone fixe, imprimante, armoire, chaise, périphériques PC, mappe monde.

- a- Image vidéo d'un arrière plan avec objets administratifs.
- b- La même image avec introduction d'un individu.



(a) Image de référence



(b) Image N° 462

Figure 3.3 : Images vidéos de type environnement administratif.



TO REMOVE PLEASE REGISTERED WITH [HTTP://WWW.PDF-CONVERTER-SOFTWARE.COM](http://www.pdf-converter-software.com)

Ces images sont les résultats d'images vidéo où l'individu est en mouvement contrairement aux images qui vont suivre où l'individu s'arrêtant momentanément de bouger devient pour un laps de temps partie de l'arrière plan et s'affichant automatiquement une fois qu'il reprend son mouvement.



(a)



(b)

Figure 3.7 : cas de figure où l'individu s'intègre à l'arrière plan.

(a) : image extraite de la vidéo.

(b) : image résultat de l'algorithme des gaussiennes multiples.

Différents agents de la nature peuvent avoir des effets conséquents à leurs intensités. L'exemple suivant le montre.



(a)



(b)

Figure 3.8 : Image résultat des effets d'agents de la nature au moment de la prise vidéo.

(a) : image de référence extraite de la vidéo.

(b) : image résultat de l'algorithme des gaussiennes multiples.

To Remove Please Registered With <http://www.pdf-converter-software.com>

### 3.2 Résultats des algorithmes de post-traitement

#### 3.2.1 Dilatation automatique

Cette étape consiste à l'amélioration des résultats des gaussiennes multiples par le remplissage des vides situés sur le corps de l'objet de façon qu'il soit plus visible. L'exemple qui suit le montre.

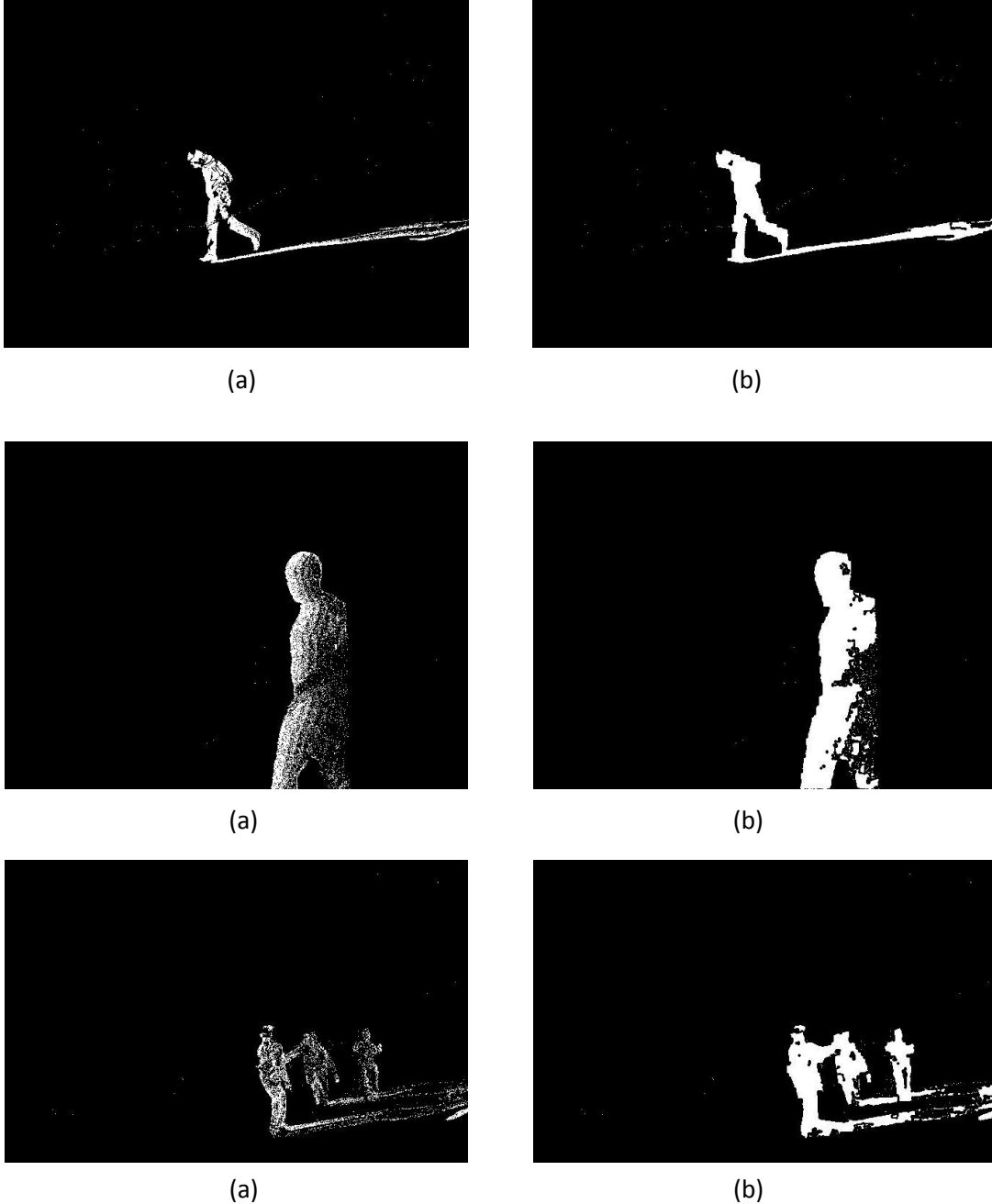


Figure 3.9 : Résultat de la dilatation automatique.

- (a) : image résultat de l'algorithme des gaussiennes multiples.
- (b) : image après la dilatation automatique.



www.pdf-converter-software.com

### 3.2.2 Élimination des points isolés

Consiste à éliminer les points inutiles à la périphérie de l'individu. L'exemple qui suit le montre.

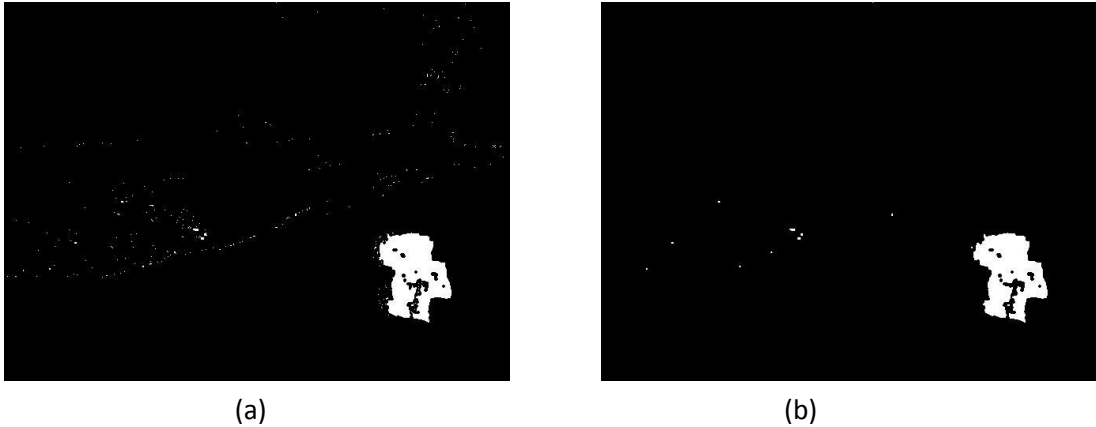


Figure 3.10 : Résultat de l'élimination des points isolés.

- (a) : image après dilatation.
- (b) : image dilatée avec élimination des points isolés.

### 3.3 Résultats de l'algorithme de suivi d'objets

Consiste à mettre toute objet en mouvement dans une boîte englobante dans l'espace et dans le temps. L'exemple suivant le démontre.

OMMO OMOO OMOO OMOO OMOO OMOO OMOO OMOO

OMMO OMOO OMOO



(a)



(b)



(a)



(b)

*Figure 3.11 : Résultat de l’algorithme de suivi d’objets.*

- (a) : image extraite de la vidéo.
- (b) : image résultat de l’algorithme de suivi d’objets.

### 3.4 Cas où la dilatation n’est pas un succès

Si la dilatation n’est pas exemplaire, le suivi d’objets ne l’est pas aussi. Les images suivantes le montre et l’individu est pris dans plusieurs boites englobantes et quelques fois une partie de son corps n’est pas prise en considération.





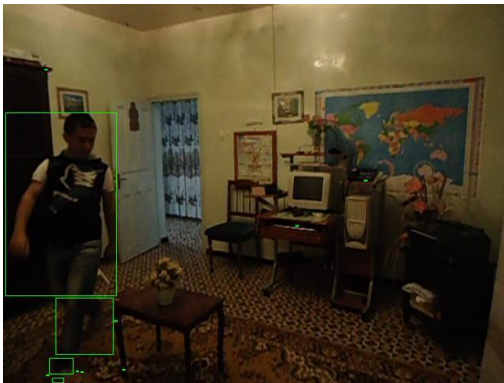
(a)

(a)



(b)

(b)



(c)

(c)

Figure 3.12 : Cas échoué de suivi d'objets.

- (a) : image extraite de la vidéo.
- (b) : image après dilatation.
- (c) : image après le suivi d'objets.

To Remove Please Registered With <http://www.pdf-converter-software.com>

## **4. Conclusion**

En conclusion, le travail que nous avons effectué nous a permis de réaliser par des algorithmes les objectifs de notre recherche qui sont la détection, la segmentation et le suivi d'objets. Les résultats sont encourageants mais nous suggérons de continuer la recherche dans ce sens pour améliorer encore plus les résultats surtout pour arriver à une dilatation optimale qui nous permis un suivi presque parfait.



## • Bibliographie

[AP, 07] : Arnaud CAPRI «Caractérisation des objets dans une image en vue d'une aide à l'interprétation et d'une compression adaptée au contenu : application aux images échographiques» thèse présentée à l'université d'Orléans pour obtenir le grade de :

Docteur de l'université d'Orléans 2007.

[FSHS, 09] : Ferjani Mohammed ,Seriai Abderrahmane ,Hennani Hakim et Seddik Annes «La Vidéosurveillance à distance» Rapport du projet TER master1,université Montpellier2 en 2009.

[GCMT, 06] : Y. Goyat, T. Chateau, L. Malaterre et L. Trassoudaine. «Vehicle trajectories evaluation by static video sensors». 9th IEEE International Conference on Intelligent Transportation Systems, 2006.

[KS, 96] :K.Tabari,S.Tagma «Compression d'images animées à très faible débit par la géométrie des fractales» PFE USTHB 1996

[MH, 97] : M.hadallah«Codage des images fixes par une méthode hybride basée sur la QV et les approximations fractales»PFE USTHB 1997.

[NT, 09] : Nguyen Thi Thuong Huyen, « Recherche d'images par le contenu avec séparation avant-plan/arrière-plan» Rapport final Juillet 2009.

[SF, 08] : M. H. Sigari et M. Fathy. «Real-time Background Modeling/Subtraction using Two-Layer Codebook Model». International MultiConference of Engineers and Computer Scientists 2008.

[TGG, 99] : N. Tronson, Y. Goyat et D. Gruyer «Comparaison de méthodes d'extraction fond/forme pour des scènes de circulation routière», Laboratoire Central des Ponts et Chaussées 1999.



To Remove Please Registered With <http://www.pdf-converter-software.com>

PDF CONVERTER SOFTWARE