

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR DE
LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DES SCIENCE ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENNE
USTHB / ALGERIE
FACULTE D'ELECTRONIQUE ET D'INFORMATIQUE



MEMOIRE

Présenté pour l'obtention du diplôme de MAGISTERE

En : INFORMATIQUE

Spécialité : Systèmes Intelligents et Ingénierie du Logicielle

Par : El Saidaoui Bassel

Sujet :

Détection d'évènements anormaux dans une scène de foule

Soutenu publiquement le 14/10/ 2010, devant le jury composé de :

Mme H. DRIAS, Professeur, à l'USTHB

Présidente

Mr S. LARABI, Professeur, à l'USTHB

Directeur de thèse

Mr H. AZZOUNE, Maître de Conférences/A, à l'USTHB

Examineur

Mr A. ABDELI, Maître de Conférences/A, à l'USTHB

Examineur

Mme S.AOUAT, Maître de Conférences/B, à l'USTHB

Invitée

Remerciements

Je tiens à remercier :

Pr. LARABI, de m'avoir encadré tout au long de ce magistère, pour ses conseils, ses remarques, sa disponibilité et surtout sa grande patience. Je le remercie aussi de m'avoir fait découvrir un domaine aussi intéressant.

Les membres de jury, d'avoir accepté de juger ce travail malgré leur emploi de temps très chargé.

Tous les enseignants de l'école doctorale d'Informatique de l'USTHB.

Un très grand merci à mon épouse.

Un très grand merci à mes parents, mes frères et sœurs.

Pour finir, je remercie tous mes collègues et mes amis.

EL-SAIDAOUI Bassel

Table des matières

Résumé	6
Abstract	7
Introduction Générale	8
1. Problématique.....	8
2. Contribution	9
3. Organisation du mémoire	9
Chapitre 1.....	10
Etat de l'art	10
.....	10
1. Introduction	10
2. Objectifs de la vidéo surveillance.....	10
3. Motivations de la surveillance automatisée.....	11
Figure1.1 : Système de vidéo surveillance avec analyse manuelle	11
Figure1.2 : Système de vidéo surveillance automatique	12
4. Evènements anormaux	12
5. Détection et suivi du mouvement	13
5.1 Détection de mouvements	13
5.2 Le Suivi de Mouvement	14
Figure1.3 : Cas où le flot optique est différent du mouvement	17
6. Détection des points d'intérêts	19
8. Détection d'événements anormaux.....	22
8.1 Types d'événements anormaux	23
8.2 Méthodes de détection d'événements anormaux : Travaux antérieurs	23
9. Conclusion	30
Chapitre 2.....	31

Conception.....	31
1. Introduction	31
2. Données du problème et Objectifs.....	31
3. Schéma général de la solution proposée	32
4. Description des traitements	33
4.1. Traitement du bas niveau.....	33
4.2 Traitement du niveau intermédiaire	36
4.3 Traitement de haut niveau	42
5. Conclusion	48
Chapitre 3.....	50
Résultats Expérimentaux.....	50
1. Données expérimentales	50
2. Application des traitements	50
3. Les résultats de détection d'évènement anormal	51
4 Conclusion	56
Conclusion générale.....	58
Annexe 1.....	60
Références	63

Table des figures

Figure 1.1	: Système de vidéo surveillance avec analyse manuelle	11
Figure 1.2	: Système de vidéo surveillance automatique.....	12
Figure 1.3	: Cas où le flot optique est différent du mouvement.....	17
Figure 1.4	: Différents types des points d'intérêts : coins, jonction en T et point de forte variation de texture.....	19
Figure 1.5	: Différentes situations considérées par le détecteur de Moravec.....	20
Figure 1.6	: Exemple de carte de direction (direction map) [IHA08].....	25
Figure 1.7	: Modélisation HMM pour les trajectoires d'objets.....	27
Figure 2.1	: Etapes de l'approche proposée.....	32
Figure 2.2	: Exemple de calcul des points d'intérêt. a) points d'intérêt à l'instant t, b) points d'intérêt à l'instant t+1.....	34
Figure 2.3	: Flot optique L-K avec pyramide d'image.....	35
Figure 2.4	: Exemple des vecteurs de mouvement d'une image.....	36
Figure 2.5	: Les vecteurs du flot optique de B (i,j) et le vecteur moyen de mouvement.....	37
Figure 2.6	: Le vecteur de référence du bloc B (i,j).....	38
Figure 2.7	: A gauche : les vecteurs de mouvement pour une image. A droite : les vecteurs moyens calculés pour la même image.....	39
Figure 2.8	: Exemple de calcul erroné de la moyenne des mouvements.....	39
Figure 2.9	: Résultat après clustérisations pour k = 3.....	41
Figure 2.10	: Carte des vecteurs d'historique.....	42
Figure 2.11	: quantification de l'écart entre les directions du vecteur moyen et le vecteur de référence.....	44
Figure 2.12	: Détection des blocs où le MCF est égale à 2,6 et 10 dans l'image.....	47
Figure 2.13	: Histogramme correspondant aux valeurs 2,6 et 10 du facteur MCF.....	48
Figure 3.1	: Evènement anormal qui commence à partir du frame 196.....	52
Figure 3.2	: Nombre d'occurrence des valeurs anormales de MCF pour les différentes frames de la vidéo.....	53
Figure 3.3	: Evènement anormal qui commence à partir du frame 157.....	54
Figure 3.4	: Nombre d'occurrence des valeurs anormales de MCF pour les différentes frames de la vidéo.....	54
Figure 3.5	: Evènement anormal qui commence à partir du frame 133.....	55
Figure 3.6	: Nombre des valeurs anormales de MCF pour les différentes frames de la vidéo.....	56

Résumé

Dans cette thèse, nous présentons une contribution pour la détection des situations anormales dans le mouvement de foule à partir d'une séquence vidéo. Notre approche se base sur l'extraction et l'analyse des vecteurs de mouvement des personnes.

Nous traitons le problème comme étant un calcul de primitives en trois niveaux : le niveau bas, le niveau intermédiaire et le niveau haut. Les caractéristiques du niveau bas et du niveau intermédiaire sont génériques et indépendantes du type d'anormalité que nous désirons détecter. Cependant, les caractéristiques du haut niveau dépendent du type d'anormalité à détecter. Dans le niveau bas nous extrayons un ensemble de points d'intérêt (POI). Par la suite, la technique du flux optique est utilisée sur ces points d'intérêt pour extraire des informations telles que la direction, la magnitude et la vitesse. Dans le traitement du niveau intermédiaire notre image est segmentée en blocs où chaque bloc contient un vecteur de direction. Nous utiliserons une méthode de clustering appelée k-means pour regrouper tous les vecteurs de chaque bloc pour obtenir une ou deux directions. Le niveau haut est constitué de mesures pour la détection d'événements anormaux. L'approche proposée a été validée sur des vidéos contenant des scènes de chute de personnes dans un escalateur d'aéroport.

Abstract

In this thesis, we present a contribution to the detection of abnormal situations in the movement of crowds from a video sequence. Our approach is based on the extraction and analysis of vectors of movement of persons.

We treat the problem as a calculation of features in three levels: low level, intermediate level and high level. The low and intermediate levels of features are generic and independent of the type of abnormality we would like to detect. However, the high level features are dependent of the type of the abnormality, we would like to detect. In the low level, we extract a set of Points of interest (POI). Optical flow techniques are used to extract information such a direction, magnitude and velocity. Our image is segmented to blocs and each bloc contains a vectors direction. We use in the intermediate level a clustering method named k-means to regroup all the vectors of each bloc in two or one direction. The high level is consists of measures for the detection of abnormal events. The proposed approach was validated on video containing scenes of people falling.

Introduction Générale

1. Problématique

Aujourd'hui, nous trouvons des applications de la vidéo surveillance partout : dans les parkings, les musées, les hôpitaux, les gares ...etc. La vidéo surveillance nous aide à observer beaucoup de positions différentes en même temps et à analyser les mouvements des personnes, des voitures et autres objets mobiles. Elle doit permettre entre autres, de détecter les événements anormaux qui se passent dans la scène à partir d'une vidéo. Lorsque le traitement n'est pas automatique, la tâche de surveillance en temps réel devient très difficile (voire impossible) à cause du nombre important de caméras (dizaines). En effet, comment un agent de sécurité pourra t-il analyser en temps réel des dizaines d'écrans de contrôle avec un risque d'erreur minimal? En plus de l'ennui ressenti en l'observant durant une longue période.

La réponse à ces deux problèmes tient en deux mots : vidéo intelligente.

La vidéo intelligente est un nouveau domaine qui traite et analyse le contenu de la vidéo. Ce terme s'exprime dans un axe de recherche assez large, qui est appliqué dans des domaines différents par exemple: la robotique, la médecine, la sécurité... . En particulier, beaucoup de recherches et de travaux dans ce domaine ont été consacrées à la vidéo surveillance. La reconnaissance des objets, des activités dans une vidéo..., sont des sujets qui sont actuellement abordés par plusieurs chercheurs.

Dans ce mémoire, nous nous intéressons à une des applications les plus importantes de la vidéo intelligente ; il s'agit de la détection des événements anormaux dans une scène de foule à partir d'une vidéo.

Plusieurs approches ont été proposées, chacune d'elles s'appuie sur un ensemble de caractéristiques en fonction de l'événement anormal traité. En outre, l'apprentissage et les processus incrémentaux et adaptifs ont été étudiés.

Cependant, les caractéristiques qu'elles utilisent sont représentées de manière spécifique, donc, la projection dans une autre application nécessite plusieurs travaux d'adaptation.

L'objectif de ce travail est d'apporter une contribution au développement d'un système qui traite et analyse la vidéo dans le but de répondre aux besoins de la sécurité.

2. Contribution

La contribution apportée dans notre approche est la détection des événements anormaux dans des scènes denses à partir des vidéos prises par un système de vidéo surveillance. Notre approche est articulée sur trois niveaux de traitement, le niveau bas, intermédiaire et haut.

Dans le niveau bas, nous extrayons un ensemble de caractéristiques telles que les points d'intérêt, le flot optique. Une méthode de classification appelée K-means est utilisée pour classer les vecteurs de mouvement en une ou deux directions dominantes. Dans le niveau intermédiaire nous calculons la variation de la direction (sens) et de la magnitude (vitesse) pour distinguer la continuité du mouvement de sa discontinuité. Dans le haut niveau, nous introduisons une mesure qui est la combinaison entre la variation de la direction et la magnitude. Cette mesure donne des valeurs caractérisant l'événement anormal.

3. Organisation du mémoire

Le reste du document est organisé de la manière suivante. Le chapitre suivant constitue un état de l'art des travaux en relation avec le thème abordé. Dans le chapitre qui le suit, nous décrirons l'approche proposée. Le dernier chapitre est consacré aux résultats de l'expérimentation. Nous terminerons par une conclusion dans laquelle nous présenterons un ensemble de perspectives.

Chapitre 1 ---

Etat de l'art

1. Introduction

De nos jours, la vidéo surveillance est un domaine de grand intérêt pour les recherches en vision par ordinateur. Cette discipline a évolué rapidement ces dernières années en vu de son importance pour assurer la sécurité des biens et des personnes. Au vu de la quantité d'informations capturée par les caméras, et au vu de cette information qui n'est autre que l'image, le problème du développement d'un système informatique pour la vidéo surveillance nécessite plusieurs techniques et particulièrement l'intelligence artificielle.

Dans ce chapitre, nous présentons d'abord les objectifs de la vidéo surveillance. Nous présenterons ensuite ce qu'est un évènement anormal et nous finirons par présenter les travaux réalisés dans ce domaine.

2. Objectifs de la vidéo surveillance

L'évolution surprenante de la technologie, particulièrement dans le domaine de la vision, vise à améliorer le quotidien des personnes dans leurs maisons, leurs lieux de travail et tout autre endroit qu'ils pourraient fréquenter. Néanmoins, la sécurité reste le domaine qui intéresse beaucoup de chercheurs. Cet intérêt porte aussi sur plusieurs secteurs tels que :

- La lutte contre le terrorisme, les crimes y compris les menaces.
- La sécurité des lieux publics, en particulier dans les réseaux de transport, les centres des villes et les établissements publics tels que les écoles, les hôpitaux et les terrains de sport.
- La gestion efficace des réseaux de transport et des services publics.

L'évolution de la surveillance a commencé dans les années 80, par le gardiennage, c'est à dire la surveillance humaine. En 1990, c'est le début de la mise en place des caméras vidéo et le commencement de la surveillance à distance, dans les routes, les centres commerciaux et les parkings. L'explosion de la vidéo surveillance a eu lieu en l'an 2000 avec l'intérêt particulier à la compréhension du mouvement du public et à la reconnaissance d'événements anormaux dans les mouvements de foules.

3. Motivations de la surveillance automatisée

Assurer la sécurité dans toutes les places publiques nécessite un nombre important de caméras et donc de vidéos qui doivent être traités et analysés automatiquement et en temps réel.

Dans les applications de vidéo surveillance qui font une analyse manuelle des séquences vidéo, les résultats ne sont pas pertinents pour les raisons suivantes :

- Le cout: une personne de contrôle pour chaque (6-10) flot vidéo.
- plus le nombre de caméras augmente, plus elles sont moins observées.
- Efficacité: la durée de vigilance d'un opérateur est d'une heure à deux heures, au-delà, son attention à la surveillance diminue progressivement.

Les figures 1.1 et 1.2 représentent un système de vidéo surveillance avec l'analyse manuelle et automatique respectivement.

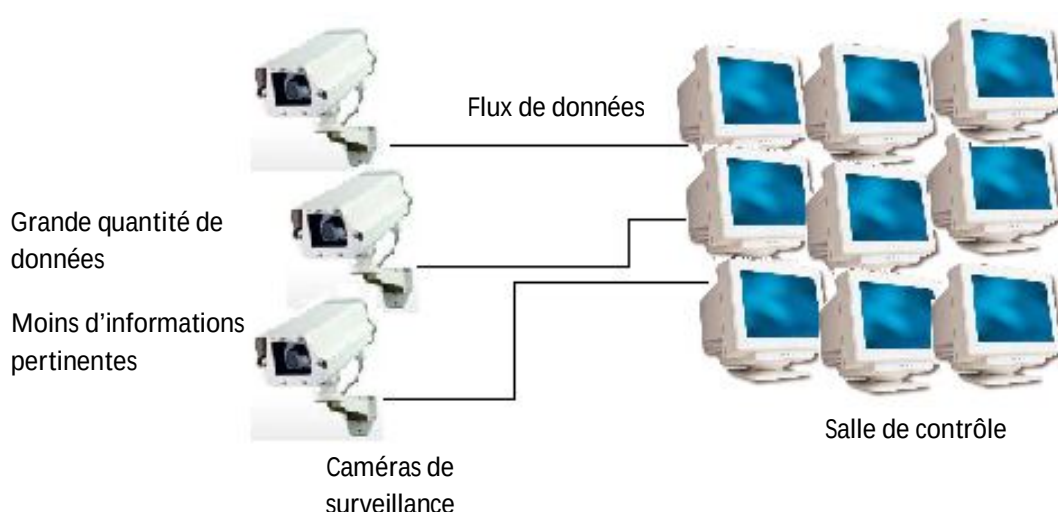


Figure1.1 : Système de vidéo surveillance avec analyse manuelle

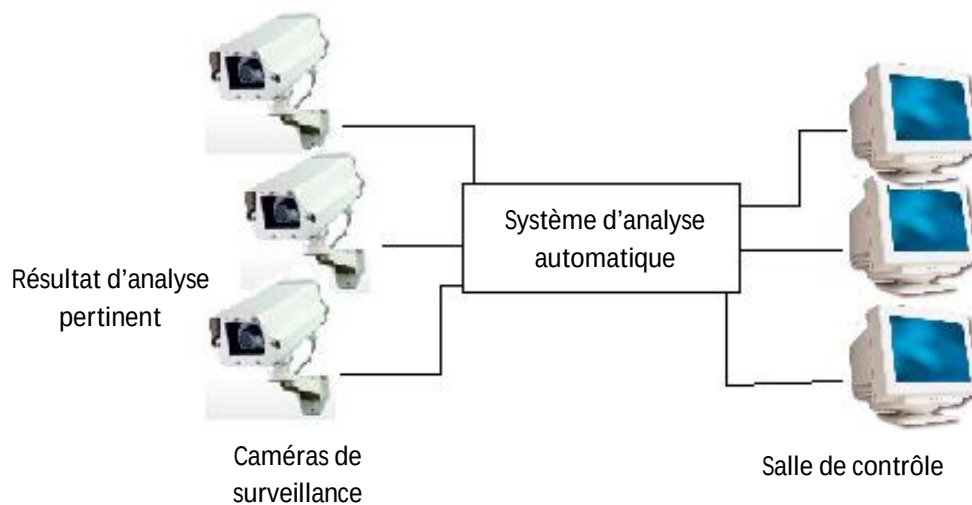


Figure1.2 : Système de vidéo surveillance automatique

Nous avons besoin donc d'un système intelligent pour traiter à temps réel la grande quantité d'information capturée par les caméras de surveillance dans le but de comprendre des événements qui se produisent dans les scènes surveillées (Détection des personnes, le suivi jusqu'à la reconnaissance du comportement ou plus précisément, la reconnaissance du comportement complexe).

4. Evènements anormaux

L'intérêt de la vidéo surveillance est de pouvoir détecter des comportements non habituels dans une scène. Ce comportement peut être celui d'un individu (par exemple, la fraude, les graffitis, le vandalisme, les attaques de banque) ; ou celui de petits groupes (par exemple, les bagarres). Il y a aussi les comportements dans des scènes de foule, par exemple l'encombrement, les bousculades, les fuites, le dépôt d'une valise, la panique ou encore les collisions entre des personnes ou des voitures.

Tous ces comportements sont classés comme des événements anormaux ne se produisant que rarement. La détection automatique de ces événements doit se faire dans un temps très court pour que l'intervention des agents sur la scène se fasse dans les délais.

Dans ce mémoire, nous nous intéressons aux comportements et aux événements anormaux dans une scène avec mouvement de foule.

Mais, pour bien détecter les événements anormaux, nous devons d'abord détecter le mouvement dans une séquence d'image, puis effectuer l'analyse nécessaire pour connaître le type d'évènement. Pour cela, nous allons présenter dans le paragraphe suivant quelques définitions concernant la détection et le suivi du mouvement et ses techniques, ainsi que les traitements nécessaires à effectuer sur ces mouvements pour détecter les événements anormaux.

5. Détection et suivi du mouvement

5.1 Détection de mouvements

La détection du mouvement dans une séquence d'images est la première étape effectuée dans chaque système de vidéo surveillance. Elle consiste à séparer, dans chaque image de la séquence, les zones en mouvement des zones statiques. La technique utilisée est le **soustraction de fond**. Dans cette technique chaque pixel doit être étiqueté par un identifiant binaire fixe/mobile. Cette différence est effectuée entre un modèle estimé de fond (image de référence), et l'image actuelle.

En d'autre terme, soient : $\mathbf{B}_t$ le modèle de fond estimé, $\mathbf{I}_t$ l'image courante ou l'image d'entrée et un seuil $\mathbf{T}$. Le pixel sera classé comme étant du premier plan si :

$$|\mathbf{I}_{t(x,y)} - \mathbf{B}_{t(x,y)}| > \mathbf{T} \quad (1.1)$$

Les techniques de soustraction de fond doivent être adaptables aux changements réels influencés par des conditions d'environnement différentes, telles que le changement d'éclairage. Pour cela, plusieurs méthodes adaptives ont été proposées pour estimer le modèle de fond et effectuer sa mise à jour en respectant l'image actuelle.

La méthode appelée *Running Average* [WRE97] propose de modéliser le fond dans chaque pixel. Le modèle est basé sur une fonction de densité de probabilité gaussienne pour les n dernières valeurs du pixel. Afin d'éviter de construire la fonction de la densité de probabilité PDF (Probabilite Densite Function) à partir du zéro, la méthode *Temporal Medain Filter* [VEL01] propose d'utiliser la valeur médiane des n dernières frames comme un modèle de

fond. Pour plus de détails concernant ces méthodes, le lecteur peut les trouver dans [PIC04], [CHE04], [HAL05].

Ces deux méthodes ont deux inconvénients :

1. Les objets de premier plan qui ont la même couleur que le fond ne peuvent pas être détectés par le seuil,
2. Ces méthodes s'adaptent lentement aux petits changements.

Pour faire face à ces problèmes, plusieurs méthodes ont été proposées. La méthode la plus utilisée dans la littérature, s'appelle *Mixture de Gaussienne* (Gaussian Mixture Model ou GMM) [STA99]. Chaque pixel de l'image est modélisé par un mélange de Gaussiennes (de 3 à 5), qui permet de déterminer quelle Gaussienne correspond aux couleurs de fond. Les pixels dont les valeurs ne correspondent pas aux distributions de fond sont considérés comme des pixels de premier plan jusqu'à leur intégration dans le modèle de fond. Pour cela, cette méthode est dite adaptative, car elle permet l'intégration de nouveaux objets et supporte les variations d'éclairage.

Une autre méthode nommée « Eigenbackgrounds » [OLI00], où la dimensionnalité de l'espace construit à partir d'un échantillon d'images est réduite en utilisant l'analyse en composantes principales (PCA). Elle propose que l'espace réduit après le PCA ne doit représenter que les parties statiques de la scène, et montre les objets en mouvement.

Cette méthode procède comme suit : pour un échantillon de N images, la moyenne de l'image de fond est calculée et les moyennes des N images sont normalisées et rangées dans une matrice A . La matrice de covariance est calculée avec la méthode de vecteurs propres. Les M grandes valeurs propres sont retenues et constituent le modèle de fond de la scène. Toute nouvelle image I_m est projetée sur l'espace fractionné par les M vecteurs propres et une image I_m' est reconstituée en utilisant les coefficients de la projection et les vecteurs propres. Le calcul de la différence $I_m - I_m'$ permet de détecter tout mouvement.

5.2 Le Suivi de Mouvement

Le suivi d'objets dans une séquence d'images occupe une place prépondérante dans plusieurs domaines en relation avec la vision par ordinateur : surveillance, robotique, ... etc.

Deux contraintes se posent pour développer un algorithme de suivi de mouvement robuste :

- la première est la qualité du suivi
- la deuxième est l'aspect temps réel exprimé via la rapidité et la complexité de l'algorithme.

Dans la littérature, plusieurs approches ont été proposées pour répondre à ces deux exigences. La plupart tendent à décomposer le problème de suivi en deux phases : phase de modélisation de l'objet à suivre et phase de suivi du vecteur rassemblant les caractéristiques extraites de la phase de modélisation.

Pour caractériser les objets, certains auteurs se sont basés sur l'histogramme de couleurs [ARA01], d'autres ont exploité l'information donnée par le contour de l'objet [PAR00]. Vu la sensibilité de ces deux techniques aux bruits et aux occultations, des études ont été amenées pour palier à ces problèmes et précisément les approches consistant à décrire l'objet à l'aide de primitives géométriques plus robustes. Nous citons les techniques se basant sur les points particuliers [SHI94], les angles et les crêtes de l'objet [THI05]. A la fin de cette phase, un vecteur est obtenu permettant d'identifier l'objet dans la scène dans la quelle il est en mouvement.

Pour suivre le déplacement de cet objet, les techniques qui ont été proposées peuvent se répartir en deux familles. Une première famille de méthodes vise à mettre en correspondance les vecteurs extraits dans chaque trame de la séquence vidéo [PIE05] comme l'utilisation de la technique du flot optique. A l'opposé d'autres techniques utilisent ces vecteurs pour vérifier la cohérence d'une prédiction faite grâce à l'ensemble des résultats des images précédentes, c'est le cas de l'utilisation du filtre de Kalman [THI05].

5.2.1 Calcul de flot optique

Plusieurs approches ont été établies pour estimer le mouvement d'un objet dans une séquence d'image, à savoir la technique classique du flot optique [XIN05].

Cette dernière consiste à détecter un mouvement apparent dans la scène à partir des variations de l'intensité dans les images. L'utilisation de cette technique prend en compte deux contraintes :

- La première est l'hypothèse que la scène est soumise à un éclairage constant.
- La deuxième est l'hypothèse que l'objet dans la scène ne subit qu'un mouvement de translation 2D.

Ces deux contraintes ne peuvent pas être présentes tout le temps dans des vidéos réelles où l'éclairage d'une scène naturelle peut varier et le mouvement de l'objet ne se limite pas à une simple translation. L'idée est alors, de retenir uniquement le déplacement maximal détecté par la technique du flot optique que peut réaliser un objet quelconque entre deux trames successives.

5.2.2 Formalisme mathématique

Considérons un pixel d'un objet en mouvement au point **A** (les coordonnées du pixel sont (x, y) , à l'instant t) avec une intensité égale à $I(x, y, t)$.

Comme le pixel va se déplacer au un point **B** dans l'image (les coordonnées du nouveau pixel sont $(x+dx, y+dy)$ à l'instant $t+dt$), ceci implique que les coordonnées du pixel (x, y) change dans le temps, mais l'intensité de ce pixel dans l'image $I(x, y, t)$ reste constante (proposition d'intensité constante) dans les deux images.

De la même manière, l'algorithme du flot optique propose que l'intensité de chaque point dans l'image pour des objets mobiles et statiques ne change pas durant le temps avec le changement des images [RIS08].

La proposition d'intensité constante est décrite comme suit :

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (1.2)$$

Utilisant le développement de la série de Taylor (sauf le premier ordre), le coté droit de l'équation ci-dessus devient :

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \dots \quad (1.3)$$

Dans cette équation, les trois points indiquent les termes de grand ordre qui seront négligés.

En remplaçant l'équation 3 dans l'équation 2, et en éliminant les termes communs, nous obtenons

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0 \quad (1.4)$$

En divisant l'équation 4 sur dt , nous obtenons :

$$\frac{\partial I dx}{\partial x dt} + \frac{\partial I dy}{\partial y dt} + \frac{\partial I}{\partial t} = 0 \quad (1.5)$$

Ceci implique:

$$\frac{\partial I dx}{\partial x dt} + \frac{\partial I dy}{\partial y dt} = -\frac{\partial I}{\partial t} \quad (1.6)$$

Qui peut être réduit à:

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y = -\frac{\partial I}{\partial t} \quad (1.7)$$

Où, V_x et V_y sont les composantes de vitesse du flot optique associées au pixel considéré, portant sur l'axe des X et l'axe des Y respectivement. Les termes $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ et $\frac{\partial I}{\partial t}$ sont les dérivées de l'image en (x, y, t) qui peuvent être écrit comme I_x , I_y et I_t . La formule précédente sera alors:

$$I_x V_x + I_y V_y = -I_t \quad (1.8)$$

Nous avons ici une équation à deux inconnus, c'est-à-dire que nous avons besoin d'autres contraintes. Pour cela, le flot optique ne correspond pas, parfois, au mouvement. Ceci est appelé le problème d'ouverture de fenêtre, qui est représenté dans la figure 1.3.

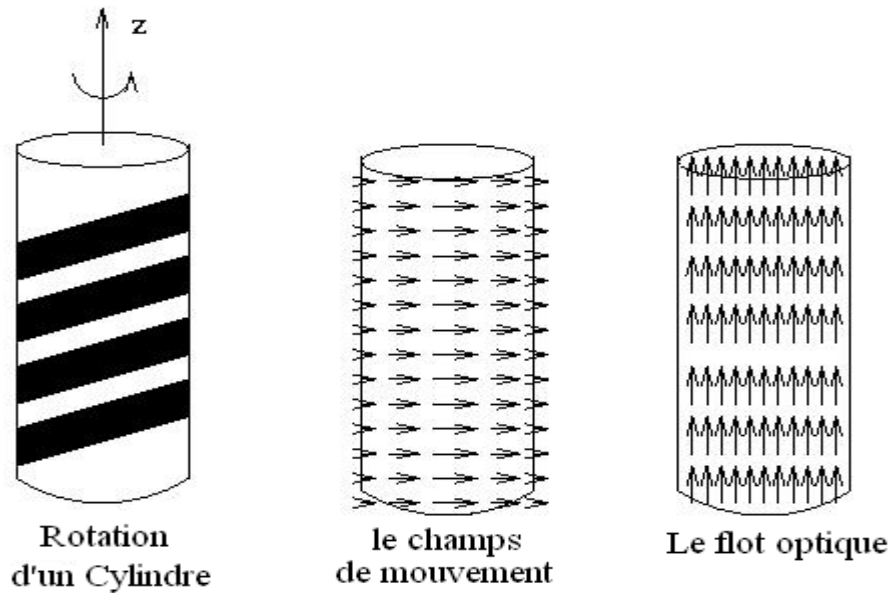


Figure1.3 : Cas où le flot optique est différent du mouvement

Il existe plusieurs méthodes pour calculer le flot optique :

- Méthodes **éparses** en flot optique [GAR08]: Ces méthodes ont besoin de spécifier à l'avance le nombre de points qui doivent être suivis. Si ces points ont des propriétés souhaitables tels que les coins alors le suivi sera robuste et fiable. Pour plusieurs applications, le coût de calcul pour ces techniques est beaucoup moins que les techniques denses (voir le point suivant). Ces méthodes nous permettent de suivre le mouvement plus rapidement. Parmi elles nous citons:
 1. La méthode de *Lukas et Kanade* [LUC81]: c'est une méthode différentielle de flot optique entre deux images. Elle introduit une deuxième hypothèse qui indique que, les pixels voisins du pixel considéré ont la même vitesse.
 2. La méthode de *Lukas et Kanade avec pyramide d'image* : qui sera utilisée dans notre travail. Elle permet de calculer le flot optique en utilisant deux images comme la méthode précédente, mais à différentes résolutions. (elle sera développée plus loin)
- Méthodes **dense** de flot optique : les méthodes de ce type sont rarement utilisées. Car elles ne supportent pas la mise en correspondance au sein d'une pyramide d'images et ne permettent pas de suivre les grands mouvements. Parmi elles nous trouvons :
 1. La méthode de *Horn et Schunck* [HOR81]: c'est une méthode qui introduit la contrainte de lissage pour résoudre le problème d'ouverture de fenêtre et elle est très sensible au bruit.
 2. La méthode *Block-matching* [HUA95; BEA95]: cette méthode tente de diviser l'image précédente et l'image actuelle en blocs (ce sont des blocs carrés qui contiennent un nombre de pixels) et calcule ensuite le mouvement de ces blocs.

Les méthodes de détection de mouvement citées ci-dessus sont utilisées dans le cas d'une foule à faible densité. Dans ce cas, nous utilisons la Mixture Gaussienne pour détecter le premier plan.

Pour le calcul du flot optique, nous utiliserons les points d'intérêts qui feront l'objet de la section suivante.

6. Détection des points d'intérêts

La détection de points d'intérêts (ou *coins*) est, au même titre que la détection de contours, une étape préliminaire pour de nombreux processus de vision par ordinateur. Les points d'intérêts, dans une image, correspondent à des doubles discontinuités de la fonction d'intensités. Ce sont par exemple : les coins, les jonctions en T ou les points de fortes variations de texture (Voir figure1.4).



Figure1.4 : Différents types des points d'intérêts : coins, jonction en T et point de forte variation de texture

Les avantages d'utiliser les points d'intérêts sont :

- Ils représentent une source d'information plus fiable que les contours car plus de contraintes sur la fonction d'intensité.
- Ils sont robustes aux occultations (soit occultés complètement, soit visibles).
- Pas d'opération de chaînage.
- Présents dans une grande majorité d'images.

De nombreuses méthodes ont été proposées pour détecter des points d'intérêts. Elles peuvent être classées en trois catégories:

1. **Approches contours** : l'idée est de détecter les contours dans une image dans un premier temps. Les points d'intérêts sont ensuite extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.
2. **Approches à base de modèles** : les points d'intérêts sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique de cette fonction des points d'intérêts considérés.
3. **Approches intensité** : l'idée cette fois-ci est de regarder directement la fonction d'intensité dans les images pour en extraire directement les points de discontinuités. Cette catégorie est généralement la plus utilisée.

Dans cette dernière catégorie, nous trouvons l'algorithme SIFT [LOW04] (Scale Invariant Feature Transform) proposé pour détecter et décrire des zones d'intérêts (local features) dans une image. A noter qu'il s'agit ici non seulement de détecter mais aussi de caractériser par des valeurs, pour pouvoir reconnaître (mettre en correspondance) ces zones ou points d'intérêts dans d'autres images de la même scène. Cet algorithme a eu un succès très important au sein de la communauté vision, mais aussi en dehors de la communauté, et de nombreuses adaptations existent.

L'idée générale de SIFT est de trouver des *primitives* qui sont invariantes à plusieurs transformations : rotation, échelle, illumination et changements mineurs du point de vue.

Un second algorithme, le *Détecteur de Moravec* [MOR97] a pour idée de considérer le voisinage d'un pixel (une fenêtre) et de déterminer les changements moyens de l'intensité dans le voisinage considéré lorsque la fenêtre se déplace dans diverses directions. Plus précisément, on considère la fonction:

$$E(x, y) = \sum_{u,v} w(u, v) |I(x + u, y + v) - I(u, v)|^2 \quad (1.9)$$

Où:

- w spécifie la fenêtre (voisinage) considérée (égale à 1 à l'intérieur de la fenêtre et 0 à l'extérieur).
- $I(u, v)$ est l'intensité au pixel (u, v) ;
- $E(x, y)$ représente la moyenne du changement d'intensité lorsque la fenêtre est déplacée de (x, y) .

La figure ci dessous représente les différentes situations considérées par le détecteur de Moravec :

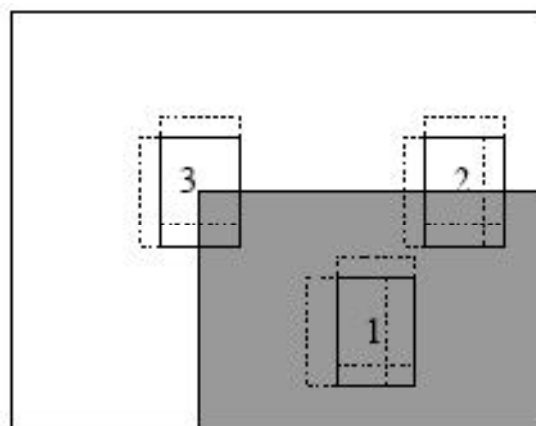


Figure1.5 : Différentes situations considérées par le détecteur de Moravec

En conséquence, le principe du détecteur de Moravec est de rechercher les maximaux locaux de la valeur minimale de E en chaque pixel (au dessus d'un certain seuil).

Le détecteur de Moravec fonctionne dans un contexte limité. Il souffre en effet de nombreuses limitations. *Harris et Stephen* [HAR88] ont en identifié certaines et en les corrigeant, ont déduit un détecteur de coins très populaire : le *détecteur de Harris*. Les limitations du détecteur de Moravec prises en compte sont :

1. La réponse du détecteur est anisotropique en raison du caractère discret des directions de changement que l'on peut effectuer (pas de 45 degrés). Pour améliorer cet aspect, il suffit de considérer le développement de Taylor de la fonction d'intensité I au voisinage du pixel (u, v) .
2. La réponse du détecteur de Moravec est bruitée en raison du voisinage considéré. Le filtre w utilisé est en effet binaire (valeur 0 ou 1) et il est appliqué sur un voisinage rectangulaire. Pour améliorer cela, Harris et Stephen propose d'utiliser un filtre Gaussien.
3. Enfin, le détecteur de Moravec répond de manière trop forte aux contours en raison du fait que seul le minimum de E est pris en compte en chaque pixel. Pour prendre en compte le comportement général de la fonction E localement, nous écrivons [MOS1] :

$$E(x, y) = (x, y) \cdot M \cdot (x, y)^t \quad (1.10)$$

Où

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (1.11)$$

La matrice M caractérise le comportement local de la fonction E , les valeurs propres de cette matrice correspondent en effet aux courbures principales associées à E :

- Si les deux courbures ont de petites valeurs, alors la région considérée a une intensité approximativement constante.
- Si une des courbures a une grande valeur et que l'autre a une petite valeur alors la région contient un contour.
- Si les deux courbures ont une grande valeur alors l'intensité varie fortement dans toutes les directions, ce qui caractérise un coin.

La définition originale de Harris nécessite de prendre le déterminant de M , et de le soustraire de la trace de M (avec quelque coefficients), puis comparer la différence à un seuil donné :

$$R = \text{Det}(M) - k\text{Trace}(M)^2 \quad (1.12)$$

Avec :

$$Det(M) = AB - C^2$$

Et

$$Trace(M) = A + B$$

Les valeurs de R sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

La méthode de *Shi et Tomasi* [SHI94] [GAR08] repose fortement sur la méthode de Harris. Ils ont trouvé que si la plus petite des deux valeurs propres était supérieure à un seuil minimal alors nous avons de bons coins.

8. Détection d'événements anormaux

Elle consiste à détecter tout événement qui n'a pas l'habitude de se produire en scène et qui peut avoir des conséquences dangereuses sur les biens ou les personnes. Cette détection est effectuée en utilisant une connaissance a priori sur les événements normaux qui se produisent tous les jours et dans différentes conditions (nuit/jour, pluie/soleil, interne/externe,...etc) et d'estimer les déviations entre les événements normaux (historique) et les événements observés. Ces déviations lorsqu'elles sont importantes, elles sont déclarées comme des événements anormaux [IHA08].

La détection d'évènement anormal est constituée de trois étapes : la première permet d'extraire toutes les caractéristiques des objets mobiles et immobiles telles que :

- les points d'intérêt, en utilisant la technique du flot optique dans le cas d'une foule dense.
- ou bien l'avant plan (foreground) en utilisant l'une des méthodes de soustraction de fond dans le cas de foule à faible densité (Mixture gaussienne est la plus utilisée [STA99]).

La deuxième étape, consiste à effectuer quelques opérations sur les caractéristiques extraites, qui donneront plus d'information sur le mouvement détecté à l'étape précédente. Ces informations concernent la vitesse, la trajectoire, la direction, le centre de gravité du blob,...etc.

Dans la troisième étape, les événements anormaux sont détectés en utilisant les résultats de l'étape précédente et une modélisation des événements normaux. Dans le cas où le décalage

(déviation entre le modèle d'événements normaux et l'image observée) est important, la scène est déclarée présentant un événement anormal.

8.1 Types d'événements anormaux

La question qui se pose, est quels sont les événements anormaux à détecter ?

Dans les places surveillées comme les hôpitaux, les écoles, les services publics, ...etc. l'apparition d'événements anormaux est fréquente et diverse. Parmi eux, il existe des événements qui ont une grande probabilité d'apparaître ou une grande possibilité de se produire, comme :

- événement de fuite (escape event) : comme le cas d'un vol ou d'une poursuite.
- événement de bousculade (collapse event) : par exemple, bousculade lors de la sortie du stade ou d'un escalator.
- événement d'abandon d'objet.
- événement de panique : fuite dans tout les sens.
- événement de jet d'objet : par exemple, la surveillance dans les terrains de sport.

Les recherches effectuées dans le domaine de détection d'événements anormaux sont vastes, et évoluent très rapidement. Il existe plusieurs méthodes de détections d'évènements anormaux. Nous allons présenter, ci-dessous, quelques unes d'elles.

8.2 Méthodes de détection d'événements anormaux : Travaux antérieurs

Il existe deux catégories de travaux de recherche. La première catégorie s'intéresse à l'analyse du flot du mouvement de foule, alors que la seconde s'intéresse à la détection des événements anormaux dans une foule.

Les travaux de la première catégorie, estime la densité de la foule [RAH06] [MAR97] [HUA04] [LIN01]. Ces méthodes sont basées sur la texture et le taux de mouvement et développe des analyses statiques pour la surveillance de la foule, mais ne détecte pas les situations anormales.

Les travaux de la deuxième catégorie détectent les événements anormaux dans la foule. L'idée générale consiste à modéliser le comportement normal, ensuite d'estimer la déviation

entre le model du comportement normal et le comportement observé. Ces déviations sont déclarées comme anormales.

Le principe de ces approches est d'exploiter le fait que les données du comportement normal sont généralement disponibles, et les données du comportement anormal sont moins disponibles.

Dans cette catégorie, nous retrouvons la méthode de [IHA08] où les auteurs ont proposé une approche de détection qui commence par l'utilisation d'une carte du mouvement dans la scène (heat map) obtenu à partir des vidéos historiques de la scène. Cette dernière permet d'indiquer les régions 'chaudes' où il y a une activité intense en mouvement. Cette carte est utilisée comme un masque pour traiter seulement les régions concernées appelées aussi régions d'intérêt. Ce masque est extrait pour les différents contextes tels que le jour/nuit, intérieur/extérieur,...etc.

Les auteurs procèdent par la suite à la détection des points d'intérêt uniquement dans les régions sélectionnées et utilisent la technique du flot optique [GIU09] pour suivre ces points durant les frames.

Pour chaque paire de frames, un ensemble de vecteurs de mouvement est calculé:

$$V = \{V1 \dots Vn | Vi = (Xi, Yi, Ai, Mi)\} \quad (1.13)$$

Où

Xi et **Yi** sont les coordonnées des vecteurs du point d'intérêt i .

Ai est la direction du mouvement.

Mi est la distance entre le point i et le point $i+1$.

Après avoir éliminé les vecteurs statiques et les vecteurs dus au bruit, ils ont divisé l'image en un nombre de bloc de taille $n \times m$. Ils ont calculé pour chaque bloc le vecteur moyen, qui représente le mouvement de ce bloc, dans le but d'obtenir une carte appelée carte de direction (Direction map) (voir figure 1.6).

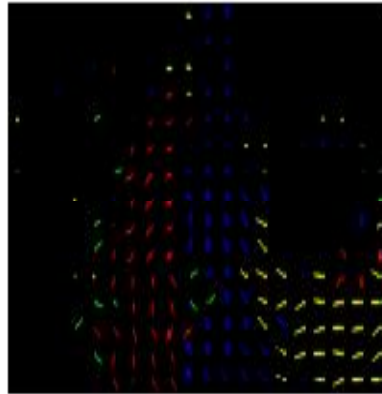


Figure 1.6 : Exemple de carte de direction (direction map) [IHA08]

Pour différencier l'anormal du normal ; ils ont introduit certaines mesures qui permettent de les distinguer. Ces mesures sont :

- Mesure du taux de mouvement qui augmente avec la densité de la foule.
- La variation de direction qui augmente dans le cas de mouvement anormal.
- La variation de magnitude qui augmente dans le cas d'un événement anormal.
- Histogramme de direction qui indique la tendance de direction, c'est-à-dire le nombre de vecteurs dans un angle donné.

La mesure proposée pour quantifier l'anormalité qui se produit dans une scène est le produit scalaire de ces différentes mesures qui augmentent dans le cas anormal et diminue dans le cas normal.

Le principal avantage de cette méthode est qu'elle ne nécessite pas une grande quantité d'information pour effectuer l'apprentissage et détecter tout les événements où la variation est importante. Dans cette méthode la sélection du seuil et la région d'intérêt doivent être rapide pour obtenir une meilleure détection.

L'approche proposée par [AND06] consiste à détecter les événements d'urgence dans les scènes de foules. Cette méthode détecte le mouvement anormal à l'aide d'une modélisation du mouvement normal et le calcul de la déviation. Pour cela, l'étape de modélisation est constituée de quatre phases : la première phase est le prétraitement qui consiste à obtenir un modèle de fond. Ce dernier produit un masque avec les objets du premier plan dans chaque frame. En parallèle avec l'extraction du premier plan ; ils calculent le flot optique sur une séquence lissée avec un filtre Gaussien pour réduire le bruit d'acquisition. La combinaison du flot optique avec le masque de premier plan permet de ne prendre en considération que les vecteurs des objets de premier plan et de réduire le bruit d'observation. Toutes les valeurs à l'extérieur du masque de premier plan sont mises à zéro pour caractériser la région statique.

La deuxième phase est la phase d'extraction de prototype. Ils commencent par diviser la séquence en segments. Chaque segment est constitué d'un nombre de frame et chaque frame est constitué de P vecteurs de flot optique. Par la suite, ils ont effectué l'analyse en composante principale (PCA) [JON09] sur chaque frame. Les vecteurs propres avec les plus grandes valeurs propres sont choisis pour former une base de projection qui réduit la dimensionnalité de $2 \times P$ au nombre de vecteurs propres choisis. Le résultat est un ensemble de vecteurs Wn du $n^{ième}$ segment

$$Wn = \{Wn1, \dots, WnT\} \quad (1.14)$$

Où Wnt est un vecteur représentant la projection de la $t^{ième}$ frame dans le $n^{ième}$ segment sur des valeurs propres sélectionnées.

La troisième étape consiste à extraire automatiquement le modèle qui nécessite le montage d'un HMM (Hidden Markov Model) pour chaque segment de vidéo. Pour trouver le nombre optimal de modèles qui représentent le mouvement normal, il effectue le regroupement spectral sur la matrice de similarité calculée en utilisant les vraisemblances inter-segment [LAW89]. Pour cela, ils ont formé avec les vecteurs de chaque frame une Mixture de Gaussien du modèle de Markov caché qui va donner un ensemble B_n de modèles (un modèle pour chaque segment), $n = 1 \dots N$. La mesure de similarité S_{ij} entre les segments i et j de la séquence est définie comme suit :

$$S_{ij} = \frac{1}{2} \{L(W_j|B_i) + L(W_i|B_j)\} \quad (1.15)$$

Où L est le Log-Vraisemblance du modèle.

Les valeurs de similarité entre les segments forment la matrice de similarité S . Cette matrice permet de trouver automatiquement le nombre de groupes dans la vidéo.

Après le regroupement spectral, les segments de vidéo sont regroupés en un nombre de classe K . Tous les échantillons W_n dans chaque classe permettent de former un nouveaux HMM par classe M_k . Le modèle final de la séquence a la forme :

$$L(W|M) = \sum_{k=1}^K \frac{N_k}{N} L(W|M_k) \quad (1.16)$$

La détection de l'anormal est basée sur la comparaison de la vraisemblance de l'observation actuelle donnée par la banque de modèles et le seuil de détection. L'observation du $n^{ième}$ segment W_n^o est considérée anormal si :

$$L(W_n^o|M) < Th_{Ab} \quad (1.17)$$

Les résultats sur des données réelles montrent l'efficacité dans la modélisation de divers comportements qui apparaissent dans les foules normales. La méthode trouve une difficulté dans le scénario où une seule personne tombe, ce qui donne un résultat moins fiable.

D'autres types d'approches de détection des événements anormaux existent basées sur les trajectoires des objets dans la scène. Nous retrouvons celles proposées par [JIA07] et [HAN07]. Dans [JIA07], les événements sont représentés comme étant les trajectoires des objets dans certains frames qui peuvent être modélisés par des HMM. La figure 1.7.a montre deux trajectoires de piétons (ligne blanche) extraites à partir d'une vidéo surveillance d'une intersection de rues.



Figure 1.7: Modélisation HMM pour les trajectoires d'objets

Soit le vecteur caractéristique $\{(x_1, y_1), \{(x_2, y_2), \dots, \{(x_T, y_T)\}$, qui représente une trajectoire 2D ; sur lequel nous appliquons cinq états HMM avec une émission Gaussienne. où :

- $\{x, y\}$ dénote les coordonnées du centre d'objet dans chaque frame et T est la longueur de la trajectoire.
- L'ellipse noire représente la moyenne et
- les croix représentent la variance de chaque état.

La détection des événements anormaux basée sur le clustering commence par une modélisation des événements normaux. Après l'apprentissage des données qui sont prises à partir des vidéos de l'historique, ces dernières sont représentées ou paramétrées par HMM. Un clustering non supervisé est effectué basé sur une certaine mesure de similarité qui est la distance d_{ij} , entre deux trajectoires i et j , modélisée par deux HMMs m_i et m_j comme suit :

$$d_{ij} = L(i|m_i) + L(j|m_j) - L(i|m_j) - L(j|m_i) \quad (1.18)$$

Où $L(i|m_j)$ dénote le log- vraisemblance de la trajectoire i en utilisant le modèle m_j pour la trajectoire j , normalisé par la taille de la trajectoire. Il est défini comme suit :

$$L(i|m_j) = \frac{1}{T_i} \log P(i|m_j) \quad (1.19)$$

Si les trajectoires i et j sont différentes, alors leur distance d_{ij} sera très grande. Si les deux trajectoires sont similaires, leur distance sera très petite et sera égale à zéro dans le cas où les deux trajectoires sont identiques. Cependant, le problème de cette mesure de similarité est le sur-apprentissage. En effet, dans la figure (1.7a), le modèle génère des paramètres très différents pour des trajectoires similaires dans la même direction. La distance de l'équation eq.18 devient très grande pour grouper les trajectoires similaires dans le même cluster. La solution est d'utiliser plus de données similaires pour former un modèle qui permet une variation plus large comme illustré dans la figure(2.7.b). Dans le cas de cette modélisation basée sur des échantillons multiples, la distance entre deux groupes de trajectoires (groupe i et j) peuvent être similaires en utilisant l'eq18, sauf pour une modification du terme de vraisemblance, c'est pourquoi nous définissons l'équation suivante :

$$L(i|m_j) = \frac{1}{N_i} \sum_r \frac{1}{T_r} \log P(i_r|m_j) \quad (1.20)$$

Où i_r dénote la $r^{\text{ième}}$ trajectoire dans le groupe i (avec une longueur égale à T_r) et N_i et le nombre de trajectoires dans le groupe i .

Pour cela, la mesure de similarité basée sur plusieurs échantillons est plus efficace que celle d'un seul échantillon. Par exemple, la distance entre les deux trajectoires de la figure1.7(a) est égale à 263.72, par contre, la figure1.7(b) représente la distance entre deux groupes, contenant chacun 20 trajectoires, qui est égale à 22.16. Pour cela, la modélisation de HMM basée sur de multiples échantillons fournit de bonnes représentations des données de trajectoires.

Cet algorithme de clustering est appelé Dynamic Heirarchical Clustering (DHC) [DUD01] [JIA07]. Chaque trajectoire forme un groupe sur lequel nous appliquons un HMM. Il y a N groupes et N HMM. Ensuite, nous calculons la distance d_{ij} entre deux groupes i et j à l'aide de l'équation 18 et l'équation 20.

Les deux groupes i et j avec la plus petite distance d_{ij} sont fusionnés dans un seul groupe si la condition suivante est satisfaite :

$$\frac{L(i|m_i) \cdot L(i|m_j)}{L(i \cup j|m_{i \cup j})} < 1 \quad (1.21)$$

Où $L(i|m_i)$ et $L(i|m_j)$ sont la vraisemblance du groupe i et j générées par les HMM appris sur les deux groupes i et j respectivement.

$L(i \cup j|m_{i \cup j})$ est la vraisemblance des échantillons des deux groupes générés par le HMM appris avec tout les échantillons.

Si les groupes ne peuvent pas être fusionnés alors le processus de clustering est terminé. Sinon on exécute l'algorithme avec $N-1$ groupe.

Les échantillons qui sont regroupés incorrectement dus au sur-apprentissage sont corrigés graduellement durant le processus itératif.

Le processus de clustering termine avec quelques groupes de données. Ces groupes contiennent un grand nombre d'échantillons où les HMMs sont appris sur chacun d'eux. Ces HMMs, dénotés par $\{m_k\}$ ($k = 1, 2, \dots$), sont les modèles des événements normaux.

Basée sur les modèles des groupes normaux, la détection des événements anormaux peut être effectuée sur de nouvelles vidéos de données. Si nous avons une nouvelle trajectoire i , la vraisemblance de la trajectoire i observée avec n'importe quelle HMM des événements normaux m_k , est dénotée par $L(i|m_k)$. Si le maximum de vraisemblance est moins qu'un seuil donné alors la trajectoire est détectée comme anormale, avec:

$$\max_k \{L(i|m_k)\} < T_{hA} \quad (1.22)$$

Où T_{hA} est un seuil.

[ALI07] utilise une **Lagrangian Particle Dynamics** pour détecter l'instabilité du flots, cette méthode est efficace pour la segmentation de haute densité de la foule (Marathons, événements politiques,...).

Dans la même catégorie, mais pour une scène avec une petite foule, les auteurs dans [CUP04] proposent un système de surveillance visuel qui observe les objets en mouvement dans un site, effectue l'apprentissage à partir de ces observations, et détecte les événements anormaux dans le site qui ne correspond pas aux mêmes modèles en utilisant une classification hiérarchique.

Aussi, les auteurs de [BOI07] posent le problème de détection de l'irrégularité dans les données visuelles comme un processus de construction d'un puzzle : ils essaient de composer la nouvelle image en utilisant des pièces de données (pièces du puzzle) extraites à partir des exemples visuels précédents (base de données). Les régions, dont les données observées peuvent être composées en utilisant de grands fragments contigus de données à partir de base de données, sont considérées très similaires. Par contre, les régions dont les données observées ne peuvent pas être composées à partir de base de données (ou peuvent être composées, mais uniquement en utilisant de petits fragments), sont vues comme non similaires ou suspects.

9. Conclusion

Dans ce chapitre nous avons rappelé l'intérêt de la vidéo surveillance et la nécessité d'avoir un système automatisé pour gérer les données traitées. Nous avons donné aussi un aperçu sur ce qu'est la détection et le suivi de mouvement. Nous avons introduit les événements anormaux et présenté un état de l'art non exhaustif sur les méthodes de détection des événements anormaux.

La solution que nous proposons s'inspire de celle proposée dans [IHA08]. Au lieu de considérer une seule direction pour un bloc, nous proposons une classification des vecteurs de mouvements en groupes. Dans le chapitre suivant, nous allons présenter notre contribution qui permet de détecter des événements anormaux de types bousculade à la sortie d'un escalateur.

Chapitre 2

Conception

1. Introduction

Dans le chapitre précédent, nous avons présenté les principaux travaux publiés dans le domaine de la vidéo surveillance. Nous nous sommes particulièrement intéressés à l'analyse de vidéos en vu d'interpréter les mouvements des objets mobiles (humains, voitures, ..). Plusieurs approches ont été consacrées à la détection des évènements anormaux pour différentes densités de foule.

Dans le même contexte, nous nous sommes intéressés à la détection de mouvements anormaux. Nous décrivons ci-après l'approche développée en présentant au début notre organisation des traitements pour l'analyse du problème posé. Nous détaillerons par la suite chacun des niveaux de cette organisation.

2. Données du problème et Objectifs

Nous disposons d'un ordinateur qui reçoit en temps réel les images (vidéo) prises par une caméra d'une scène fréquentée par une foule tout au long de la journée.

Nous supposons aussi qu'un archive de vidéos est disponible pour une période donnée dans laquelle il n'y a pas eu de mouvement de foule anormal et que les agents de sécurité ne sont pas intervenus. Ces vidéos permettent d'avoir un historique des trajectoires des mouvements, des vitesses, des directions et donc de pouvoir reconnaître le mouvement normal et de définir ses attributs.

Il s'agit donc, de détecter à partir des images acquises, en temps réel, des vidéos de l'archive, toute situation anormale dans la scène. En d'autres termes, tout mouvement non régulier: mouvement de panique, de fuite, de bousculade, etc.

3. Schéma général de la solution proposée

La solution au problème posé est décomposée en trois niveaux de traitements :

- Les traitements du bas niveau : Il s'agit d'extraire les caractéristiques de l'image telles que les points d'intérêt, le flot optique (les vecteurs de mouvement), les régions, les contours.
- Les traitements du niveau intermédiaire: Il s'agit de définir et de calculer des caractéristiques en se basant sur les caractéristiques du bas niveau et qui pourront être utilisées pour l'interprétation des événements de la scène tels que : la vitesse, la direction, l'accélération, l'énergie, la densité de la foule, la trajectoire, etc.
- Les traitements du haut niveau : Il s'agit de définir des mesures en se basant sur les caractéristiques du niveau intermédiaire permettant de détecter les événements anormaux.

La figure2.1 montre clairement cette décomposition.

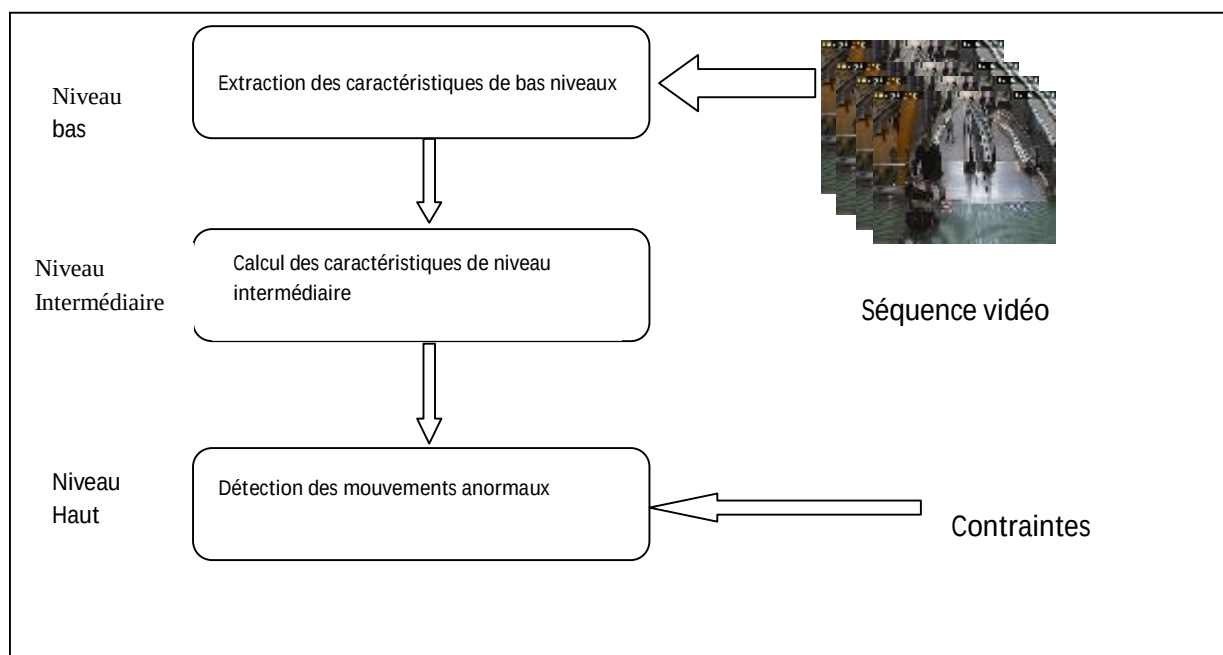


Figure2.1 : Etapes de l'approche proposée

L'objectif de cette décomposition est de pouvoir développer une bibliothèque de routines calculant les primitives des niveaux élémentaire et intermédiaire. Ces primitives seront ensuite utilisées pour mesurer l'anormalité des événements anormaux. Pour détecter chaque type d'événement anormal, un ensemble spécifique des primitives des premiers niveaux est utilisé.

Nous pouvons résumer les intérêts de cette décomposition comme suit:

- Elle est fonctionnelle, elle indique clairement qu'est ce qui doit être calculé et quand il doit être calculé.
- Elle autorise le partage de toutes les caractéristiques des deux premiers niveaux entre différentes applications.
- Elle offre la possibilité de modifier/ajouter d'autres caractéristiques à celles des deux premiers niveaux pour détecter de nouveaux types d'événements.

4. Description des traitements

4.1. Traitement du bas niveau

4.1.1 Extraction des points d'intérêt

Il s'agit d'extraire toutes les primitives sur les images d'une vidéo correspondantes aux objets de la scène. Pour rappel, les points d'intérêts dans une image correspondent à des doubles discontinuités de la fonction d'intensité. Ce sont par exemple : les coins, les jonctions en T ou les points de fortes variations de texture [MOS1].

Pour la détection des points d'intérêt, il existe plusieurs approches utilisées telles que le *détecteur de Markov*, le *détecteur de Harris* et l'algorithme *SHIFT* de David Lowe. La méthode la plus couramment utilisée est celle de Harris.

Dans notre travail, nous avons opté pour la méthode introduite par **Shi et Tomasi** [SHI94], qui donne des résultats plus satisfaisants que la méthode de Harris [MOS1]. Elle permet de localiser sur l'image un ensemble suffisant de points d'intérêt. Nous considérons aussi, que dans les scènes de vidéo surveillance, la position des caméras et les conditions d'éclairage conditionnent le nombre de points d'intérêt détectés.

La figure 2.2 illustre deux exemples de points d'intérêt calculés sur deux images différentes d'une même vidéo. Dans l'image de gauche nous détectons les points d'intérêt à l'instant t . Dans l'image de droite, nous détectons ces points à l'instant $t+1$. Le nombre de points détectés est différent à cause du changement de la texture de l'image due au mouvement ou aux conditions d'éclairage et à la position des caméras. Un même point d'intérêt peut changer

de position ou encore disparaître, puisque un mouvement donné peut changer complètement la position des objets ou des personnes qui apparaissent dans la vidéo. L'éclairage joue aussi un rôle, la texture change d'une image à une autre selon l'intensité lumineuse. Bien sûr, ces mêmes contraintes font aussi apparaître de nouveaux points d'intérêt.

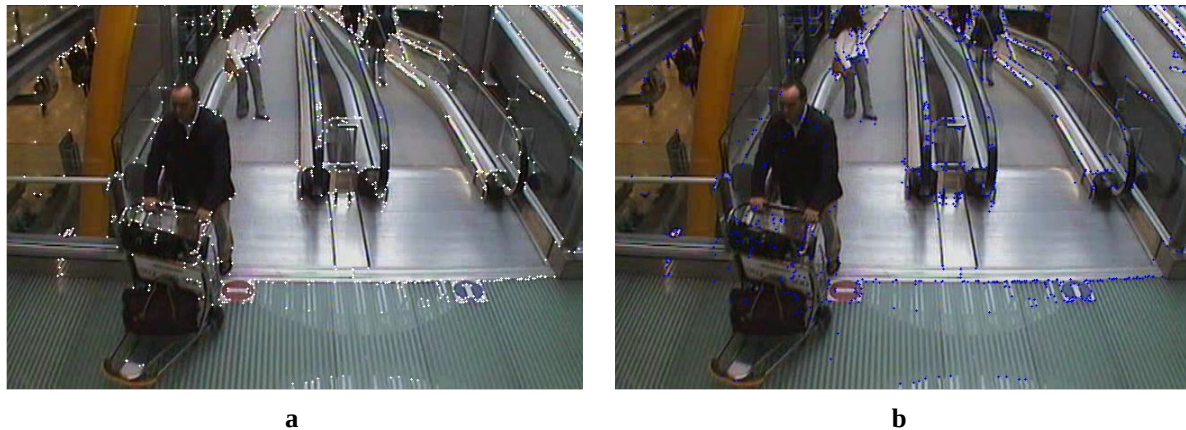


Figure 2.2 : Exemple de calcul des points d'intérêt. a) points d'intérêt à l'instant t , b) points d'intérêt à l'instant $t+1$.

Une fois l'ensemble de points d'intérêt défini, nous assurons le suivi de ces points durant l'image suivante en utilisant la technique du flot optique.

4.1.2 Calcul du flot optique

La seconde étape consiste à suivre ces points d'intérêt durant les prochaines frames en utilisant la technique du flot optique. Pour cela, nous utilisons la méthode de Lukas et Kanade avec pyramides d'image [LUC81] [SHI94]. L'algorithme de Shi et Tomasi [SHI94][GAR08] utilise les plus petites valeurs propres pour un bloc de l'image comme un critère pour assurer la sélection des caractéristiques qui peuvent être réellement suivies par le Lukas et Kanade. l'algorithme de Lukas et Kanade avec pyramides d'images [GAR08] calcule le flot optique à différentes résolutions. Cet algorithme fait la mise en correspondance automatiquement, et le résultat est un ensemble de vecteurs qui correspondent au flot optique.

La méthode de flot optique Lukas et Kanade (voir annexe 1) est une méthode à deux différentielles d'images, car elle a besoin de deux images pour fonctionner. C'est la méthode la plus populaire d'estimation de flot optique, elle a été développée par Bruce D. Lucas and Takeo Kanade [LUC81]. Elle introduit une nouvelle hypothèse qui permet d'obtenir un

système entièrement déterminé : donnant un pixel P , tous les pixels au voisinage de P sont supposés avoir la même vitesse que P .

Le principal inconvénient de la méthode Lukas et Kanade est qu'elle est fonction de la taille de la fenêtre utilisée et de la texture, en plus, elle est capable de suivre uniquement, les mouvements dans les images dont les déplacements entre deux images sont de l'ordre d'un pixel [DEV06].

Dans le but de suivre les grands mouvements, une amélioration de la méthode Lukas et Kanade avec pyramide d'image basée sur un traqueur 2-D [BOU00] est employée. Elle permet d'estimer le mouvement 3-D à un niveau plus haut (typiquement $1/8$ de la taille originale de l'image) et améliore cette estimation à chaque niveau bas ($1/4$, $1/2$ et 1 de l'image). Le suivi en utilisant les pyramides d'image permet aux grands mouvements d'être détectés par la fenêtre locale (voir figure2.3).

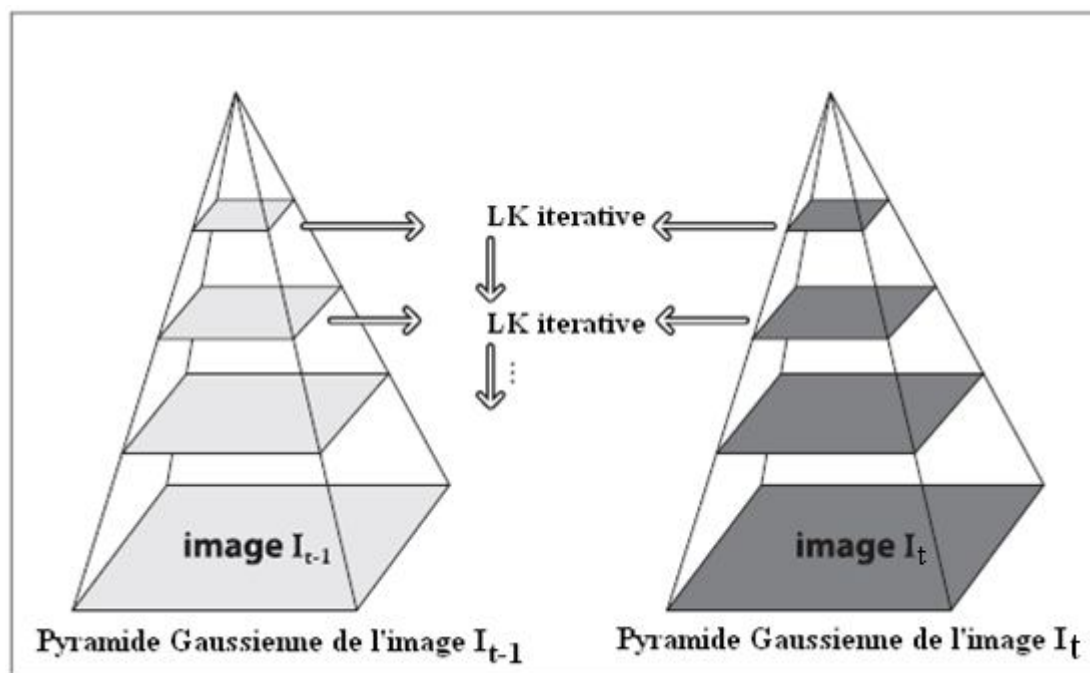


Figure2.3 : Flot optique L-K avec pyramide d'image.

Le principe de la méthode Lukas et Kanade avec pyramide d'image est de calculer le flux optique de façon pyramidale allant de la plus grande à la plus petite résolution. Par conséquent, l'algorithme peut traiter de gros flux de pixel, tout en gardant une petite fenêtre de l'intégration, et donc atteindre un niveau élevé de précision locale (de l'ordre de $0,1$ pixel) [JEA02].

Les deux méthodes de flot optique L-K et L-K avec pyramide d'image sont implémentées par la bibliothèque de vision par ordinateur OpenCV. Ces routines sont `cvCalcOpticalFlowsLK()` et `cvCalcOpticalFlowsLKPyr()`. La deuxième routine qui est utilisée dans notre travail prend en entrée deux images et un ensemble de points d'intérêt. Elle fait la mise en correspondance automatiquement, et nous retourne un ensemble de vecteurs de mouvements à deux dimensions :

$$V = \{V_1, V_2, \dots, V_N | V_i = ((X_{di}, Y_{di})(X_{ai}, Y_{ai}))\} \quad (3.1)$$

Où :

$(X_{di}, Y_{di}), (X_{ai}, Y_{ai})$: Sont les coordonnées des extrémités du vecteur V_i

Après cette étape, nous procédons à une élimination des vecteurs statiques et aux vecteurs correspondant au bruit. Les vecteurs statiques sont caractérisés par une magnitude faible alors que ceux correspondant au bruit sont reconnus par leurs très grandes magnitudes par rapport aux vecteurs de voisinage. Le résultat de cette étape est illustré par la figure2.4.



Figure2.4 : Exemple des vecteurs de mouvement d'une image

4.2 Traitement du niveau intermédiaire

Partant des primitives de bas niveau (points et vecteurs de mouvement), il s'agit de proposer un ensemble de primitives du niveau intermédiaire permettant de caractériser le mouvement de la foule dans une scène.

Dans ce travail, nous nous intéressons d'abord à la continuité du mouvement qui est caractérisée par :

- La continuité dans la direction qui dépendra de la différence d'orientation entre deux vecteurs successifs.
- La continuité dans la magnitude qui dépendra de la différence de magnitudes entre deux vecteurs successifs.

Dans ce qui suit, nous présentons notre approche pour quantifier cette caractéristique du mouvement à partir des images de vidéo.

4.2.1 Vecteur moyen du mouvement

Nous décomposons chaque image en blocs carrés de k^2 pixels. Un bloc sera noté $B(i, j)$ où i et j correspondent aux coordonnées de son premier pixel.

Après localisation des points d'intérêts et le calcul des vecteurs de mouvement, chacun des blocs $B(i, j)$ contiendra donc un ensemble de vecteurs $\overrightarrow{V_k(i, j)}$, $k = 1, \dots, n_p$ où n_p est le nombre de vecteurs de mouvement du bloc $B(i, j)$.

Le nombre de ces vecteurs et leurs directions dépend du nombre des points d'intérêt extraits et la densité de la foule et du mouvement.

Nous associons au bloc $B(i, j)$ du frame numéro n le vecteur moyen du mouvement $\overrightarrow{V^p(i, j)}$, défini comme la moyenne des vecteurs de mouvement dans ce bloc. Soit n_p le nombre de ces vecteurs, $\overrightarrow{V^p(i, j)}$ est calculé comme suit:

$$\overrightarrow{V^p(i, j)} = \frac{1}{n_p} \sum_{k=1}^{n_p} \overrightarrow{V_k(i, j)} \quad (2.2)$$

La figure 2.5 montre un exemple du vecteur moyen obtenu à chaque frame

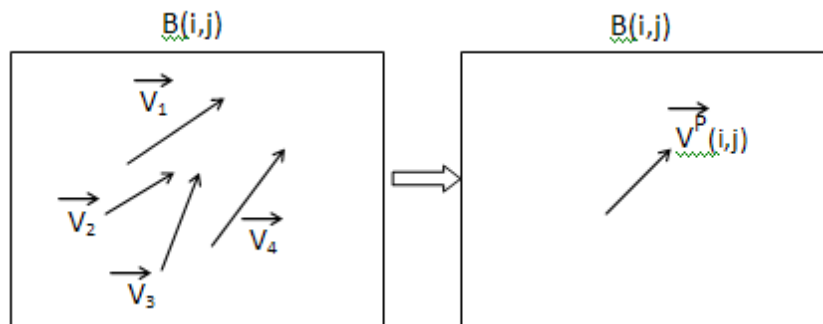


Figure2.5 : Les vecteurs du flot optique de $B(i, j)$ et le vecteur moyen de mouvement

4.2.2 Vecteur de référence

Nous définissons le vecteur de référence du mouvement du bloc B (i,j) comme la résultante des vecteurs de mouvement moyen du même bloc sur n frames de la vidéo. Ce vecteur est calculé à partir de l'historique du mouvement d'une scène donnée. Nous utiliserons donc un ensemble de vidéos qui ne contiennent pas de mouvement anormal.

Il est calculé comme suit, où **n** est le nombre de frames utilisées:

$$\overrightarrow{V^r(i,j)} = \frac{1}{n} \sum_{p=1}^n \overrightarrow{V^p(i,j)} \quad (2.3)$$

Les figures 2.6 et 2.7 illustrent ce calcul.

Par convention, le point en haut à gauche de chaque bloc est utilisé pour représenter le début de chaque vecteur moyen appartenant à ce bloc. Ceci facilitera le regroupement par la suite.

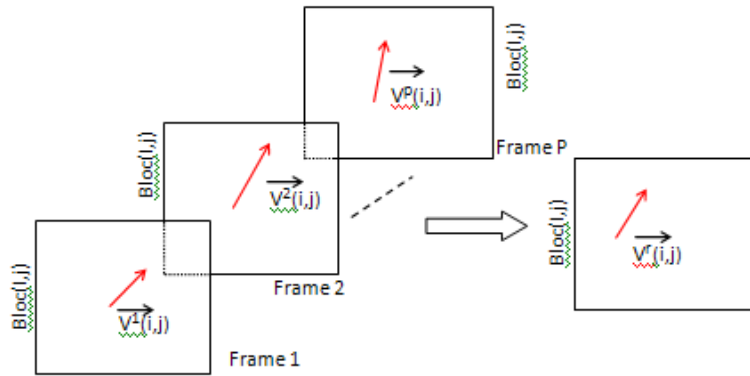


Figure2.6 : Le vecteur de référence du bloc B (i,j)

Le calcul de la résultante des vecteurs moyens de mouvement pour chaque bloc de l'image ne permet pas dans le cas général de trouver la direction du mouvement. En effet, il est possible pour un bloc donné qu'il y ait plusieurs directions (deux ou trois). Dans le cas où deux vecteurs de directions opposées appartiennent au même bloc ; leur résultante ne donne pas forcément un vecteur qui représente le mouvement. En effet, dans ce cas, nous allons obtenir un vecteur nul qui signifie qu'il n'y a pas de mouvement dans ce bloc, ce qui n'est pas vrai.

La solution à ce problème, est de regrouper les vecteurs moyens de chaque bloc en plusieurs vecteurs de référence. Pour cela, nous utiliserons une technique de regroupement appelée K-means que nous allons détailler dans la section suivante.



Figure 2.7 : A gauche : les vecteurs de mouvement pour une image. A droite : les vecteurs moyens calculés pour la même image.

Classification K-means

Après avoir calculé les vecteurs moyens de mouvement dans chacun des blocs de toutes les images acquises, nous évaluons la moyenne au niveau de chaque bloc sur plusieurs images. Si nous calculons la moyenne de plusieurs vecteurs comme précédemment, nous pourrions obtenir un faux résultat. La figure 2.8 illustre un exemple où la résultante est nulle.

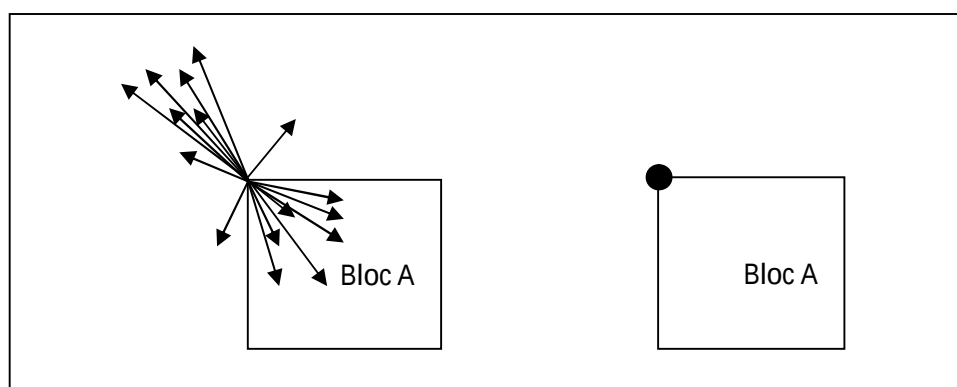


Figure2.8 : Exemple de calcul erroné de la moyenne des mouvements

Les vecteurs moyens calculés pour chaque bloc sur plusieurs images peuvent avoir des sens différents. Si nous cherchons le vecteur moyen en calculant la moyenne de tous les vecteurs alors le vecteur résultant ne correspondra pas à l'historique du mouvement du bloc. Par exemple, dans la figure 2.8, le vecteur moyen de tous les vecteurs du bloc A est un point qui indique qu'il n'y a pas de mouvement dans ce bloc (Direction et magnitude égale à zéro) alors qu'il y a différents sens du mouvement.

Comme les vecteurs peuvent avoir des sens différents, nous proposons de calculer la moyenne pour les vecteurs qui sont de même direction. Pour cela, nous proposons de grouper les vecteurs dans des classes où chaque classe contiendra les vecteurs qui ont la même direction. Par exemple dans la figure 2.8, presque la moitié des vecteurs sont dans un sens, et les autres dans le sens inverse.

Pour cela, nous introduisons dans notre travail la classification qui consiste à regrouper tous les vecteurs moyens du bloc sur plusieurs images en groupes. Chaque classe va contenir les vecteurs qui ont relativement une même direction. Pour assurer cette opération, nous avons choisi une méthode de clustérisations non supervisée appelée K-means [KAN02] [GAR08], car elle est multi-valeurs, c'est-à-dire, elle prend n'importe quelle type de données.

Donc, la méthode K-Means est un algorithme de classification qui tente de trouver les classes dans un ensemble de données. L'utilisateur fixe le nombre de classe et K-means cherche rapidement les meilleurs emplacements des centres du cluster. C'est l'une des méthodes les plus utilisées et elle est similaire à l'algorithme EM (expectation maximisation) de mixture de gaussiennes [GAR08]. Son fonctionnement est décrit comme suit:

1. Prend en entrée un ensemble de données n et un nombre de cluster k ($k < n$).
2. Aléatoirement, affecte les locations des centres des clusters.
3. Associe chaque point de donnée à la classe dont le centre est le plus proche.
4. Calcule le nouveau centre de chaque classe.
5. Retourne à l'étape 3 jusqu'à convergence (les centres des classes ne se déplacent plus).

Cet algorithme est aussi implémenté dans la bibliothèque de vision par ordinateur OpenCV. Il prend comme entrée les coordonnées des extrémités d'arrivée des vecteurs de chaque bloc, et retourne les coordonnées des k centres des clusters.

Pour que l'algorithme ne diverge pas, la routine reçoit aussi en entrée des critères d'arrêt de l'algorithme tels que le nombre d'exécution qui ne doit pas dépasser pas 10, ou que la distance entre le centre du cluster et le nouveau centre soit inférieure à 1.

Notre but est d'effectuer la moyenne des vecteurs qui appartiennent à la même classe. Comme dit précédemment, tous les vecteurs moyens d'un bloc donné ont la même extrémité de départ. Pour cela, nous allons appliquer notre algorithme de classification sur les secondes extrémités. Nous obtenons les centres des classes, regroupant la seconde extrémité des vecteurs moyens pour un bloc i sur p images, qui représentent les différentes directions pour ce bloc, comme il est montré dans la figure 2.9 (nous avons choisi $k = 3$).

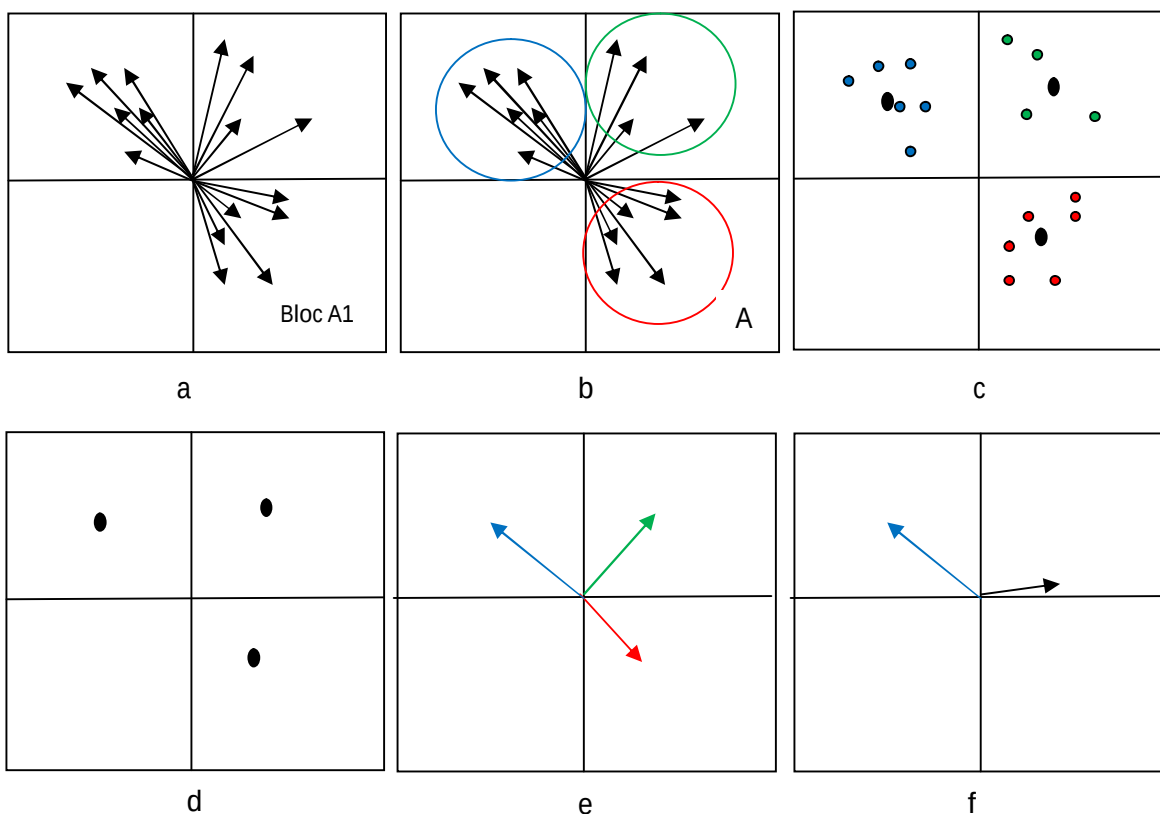


Figure 2.9: Résultat après classification pour $k = 3$.

Après avoir effectué la classification des vecteurs du mouvement en vecteurs directionnels, nous proposons de fusionner tout couple de classes si leurs vecteurs sont relativement de direction voisine. Pour cela, nous calculons d'abord l'angle entre les vecteurs moyens de chaque bloc. Si l'angle est inférieur à un angle minimal alors nous effectuons le fusionnement

qui consiste à calculer la moyenne entre ces deux vecteurs. Cette étape est refaite jusqu'à ce qu'il ne reste que des vecteurs moyens où l'angle entre eux est supérieur à l'angle minimal. Les figures 2.9.e et 2.9.f montrent ce processus.

Après que le processus de classification et de regroupement soit appliqué à tous les blocs de l'image, nous obtenons une carte appelée carte de direction. Cette dernière contiendra l'historique du mouvement durant toutes les frames de la séquence.

La figure 2.10 illustre le vecteur de référence de chaque bloc pour une scène réelle. Quelques blocs contiennent deux directions, c'est-à-dire ont deux vecteurs de référence. Cette étape est effectuée une seule fois pour une scène donnée à partir de séquences vidéos normales.

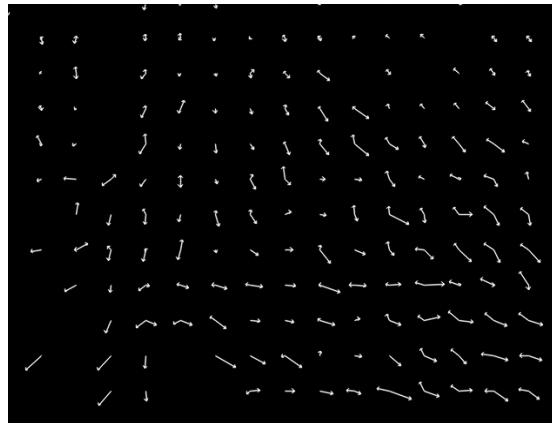


Figure 2.10: Carte des vecteurs d'historique

Chacun des vecteurs de référence calculés pour les blocs de l'image seront utilisés pour étudier la nature du mouvement pour toute autre séquence vidéo. Pour cela, nous calculons la magnitude et l'angle de chacun de ces vecteurs.

4.3 Traitement de haut niveau

Nous supposons que les vecteurs de référence sont disponibles pour chaque bloc de l'image. En utilisant un ensemble d'images acquises, nous calculons un ensemble de mesures à partir des primitives du niveau intermédiaire qui serviront pour se prononcer sur la normalité ou non du mouvement dans la scène à partir de la vidéo.

Concrètement, dans ce travail, nous introduisons deux mesures qui seront calculées sur chaque bloc de l'image pour caractériser la continuité du mouvement dans une séquence d'images vidéo. Ces deux mesures sont :

- La variation de la direction du mouvement.
- La variation de la magnitude du mouvement.

Pour chaque bloc de l'image, nous disposons de son vecteur de référence du mouvement. Nous calculons aussi son vecteur de mouvement moyen à partir de l'image courante.

Soient :

$\overrightarrow{V^r(i,j)}$ Le vecteur de référence du bloc (i, j).

$\overrightarrow{V^p(i,j)}$ Le nouveau vecteur moyen du bloc (i, j) pour la frame numéro **p**.

Ces deux vecteurs peuvent être différents en direction et en magnitude. Cette différence augmente dans le cas de mouvement anormal. Comme vu précédemment, chaque bloc peut avoir deux ou un seul vecteur moyen.

4.3.1 Variation de la direction du mouvement

Nous définissons la différence de direction **DD** entre les vecteurs $\overrightarrow{V^r(i,j)}$ et $\overrightarrow{V^p(i,j)}$ comme l'angle entre ces deux vecteurs.

$$DD = |(\overrightarrow{V^r(i,j)}, \overrightarrow{V^p(i,j)})| \quad (2.4)$$

La valeur de **DD** est comprise entre 0 et π .

Pour distinguer entre la continuité et la discontinuité dans la direction (**DCM**) nous introduisons les mesures suivantes (voir figure 2.11):

Si nous avons un seul vecteur de référence alors

$$DCM = \begin{cases} 0, & si \quad DD \leq \frac{\pi}{3} \\ 1, & si \quad \frac{\pi}{3} < DD \leq \frac{\pi}{2} \\ 2, & si \quad \frac{\pi}{2} < DD \leq \pi \end{cases} \quad (2.5)$$

Où

0, indique une continuité dans la direction du mouvement.

1, indique une légère discontinuité dans la direction du mouvement.

2, indique une importante discontinuité dans la direction du mouvement.

Dans le cas de présence de deux vecteurs de référence dans le bloc $B(i,j)$: $\overrightarrow{Vr^1(i,j)}$ et $\overrightarrow{Vr^2(i,j)}$, nous calculons leurs différences de direction DD_1 et DD_2 respectivement en utilisant l'équation 2.4, nous aurons alors (voir figure 2.11):

$$DCM = \begin{cases} 0, & \text{si } 0 < DD_1 \leq \frac{\pi}{3} \text{ ou } \frac{7\pi}{12} < DD_2 \leq \frac{11\pi}{12} \\ 1, & \text{si } -\frac{\pi}{12} < DD_1 \leq 0 \text{ ou } \frac{\pi}{3} < DD_1 \leq \frac{5\pi}{12} \text{ ou } \frac{\pi}{2} < DD_2 \leq \frac{7\pi}{12} \text{ ou } \frac{11\pi}{12} < DD_2 \leq \pi \\ 2, & \text{si } \frac{5\pi}{12} < DD_1 \leq \frac{\pi}{2} \text{ et } -\frac{\pi}{12} < DD_2 \leq -\pi \end{cases} \quad (2.6)$$

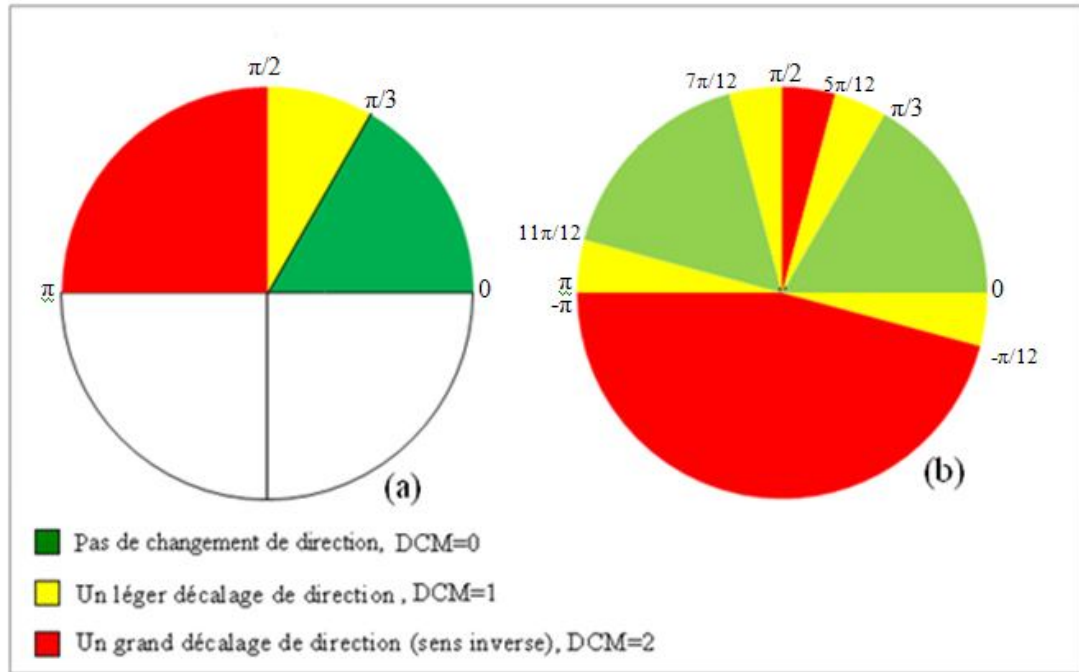


Figure 2.11 : quantification de l'écart entre les directions du vecteur moyen et le vecteur de référence

4.3.2 Variation de la magnitude du mouvement

Nous définissons la mesure de continuité de magnitude du mouvement (**MMC**) dont le but d'interpréter toute accélération ou décélération. **MMC** est définie en fonction de la différence des magnitudes des vecteurs de référence $\overrightarrow{Vr(i,j)}$ et le vecteur moyen $\overrightarrow{Vp(i,j)}$ comme suit.

$$MMC = \begin{cases} 0, & , si \, \|\overrightarrow{V^r(l,j)}\| - \|\overrightarrow{V^p(l,j)}\| > 0 \\ 4, & , si \, \|\overrightarrow{V^r(l,j)}\| - \|\overrightarrow{V^p(l,j)}\| = 0 \\ 8, & , si \, \|\overrightarrow{V^r(l,j)}\| - \|\overrightarrow{V^p(l,j)}\| < 0 \end{cases} \quad (2.7)$$

Où

0, indique une discontinuité dans la magnitude qui correspond à une décélération

4, indique une continuité dans la magnitude de mouvement

8, indique une discontinuité de magnitude de mouvement qui correspond à une accélération

4.3.3 Modélisation et détection d'événement anormal

4.3.3.1 Modélisation d'événement anormal

La continuité du mouvement est fonction des valeurs de variations de direction et de magnitude mesurées pour les frames courantes relativement aux frames des vidéos historiques.

Nous définissons le facteur de continuité de mouvement (MCF) comme étant la somme des valeurs de la continuité de direction du mouvement (DCM) et de la continuité de la magnitude du mouvement (MMC). Il s'écrit comme suit :

$$MCF = DCM + MMC \quad (2.8)$$

Donc, pour chaque frame, les variations des magnitudes et directions du mouvement sont calculées pour tous les blocs.

Les différentes valeurs du MCF sont résumées dans le tableau suivant où :

DCM \ MMC	0	4	8
0	0	4	8
1	1	5	9
2	2	6	10

Table1 : les différentes valeurs du MCF

Les valeurs du MCF sont interprétées comme suit :

- MCF = 0 continuité de direction avec une discontinuité de magnitude (décélération)
- MCF = 1 Légère discontinuité de direction et discontinuité de magnitude (décélération)
- MCF = 2 Sens opposé et discontinuité de magnitude (décélération au sens opposé)
- MCF = 4 Continuité de direction et continuité de magnitude (mouvement normal)
- MCF = 5 Légère discontinuité de direction et continuité de magnitude
- MCF = 6 Sens opposé et continuité de magnitude (mouvement normal dans le sens opposé).
- MCF = 8 Même direction et discontinuité de magnitude (accélération)
- MCF = 9 Légère discontinuité de direction et discontinuité de magnitude (accélération)
- MCF = 10 Sens opposé et discontinuité de magnitude (accélération en sens opposé)

Ces valeurs de MCF seront affectées après calcul aux blocs de l'image. Dans le cas où il n'y a pas eu de mouvement dans l'image précédente et l'image actuelle (blocs de l'arrière plan ou objets immobiles), le facteur MCF pour ces blocs est défini par défaut égal à 3.

La matrice de facteur de continuité de mouvement nous permet d'interpréter le mouvement. Par exemple, si tous les éléments de la matrice ont la valeur 4, alors les mouvements sont normaux. Si dans cette image, il y a quelques éléments qui ont la valeur 10, alors ceci signifie qu'il existe un mouvement en sens inverse avec une accélération comme le montre la figure 2.12.

Dans le cas d'un événement anormal, il est clair que des valeurs anormales de magnitude du mouvement (décélération ou accélération) et de direction du mouvement apparaissent dans un bloc B (i,j). Ces valeurs doivent aussi apparaître dans les blocs voisins durant un nombre significatif de frames.

Ainsi, dans le cas où dans le frame courant, la valeur du MCF est synonyme d'évènement anormal, un processus de suivi est activé dans le but de suivre ces valeurs anormales pour les blocs voisins pour les frames suivantes.

Dans le cas où un ensemble de blocs contigus pour un ensemble de frames successifs sont caractérisés par des valeurs anormales du MCF, un évènement anormal sera alors signalé. Cela se produit comme suit :

4.3.3.2 Détection d'évènement anormal

Nous proposons que le processus d'apprentissage soit effectué en utilisant un ensemble de vidéos normales. Nous avons vu qu'avec les valeurs du facteur de continuité de mouvement MCF, nous pouvons connaître le type de mouvement dans l'image et plus précisément le type de mouvement dans chaque bloc de l'image. Par exemple si le mouvement est normal alors MCF va avoir les valeurs 4 ou 5 dans les blocs où se situe ce mouvement.

Dans le cas d'un évènement anormal, par exemple, évènement de bousculade, la direction des vecteurs de mouvement est dans tous les sens et leur mouvement n'est pas continu. Ceci peut être détecté par un comptage des blocs de voisinage qui vérifient cette propriété pour des frames successifs. Les blocs où il y a un évènement de bousculade sont caractérisés par les valeurs du MCF 2, 6, 8, et 10.

Dans notre travail, nous détectons tous les blocs qui portent des valeurs anormales de MCF

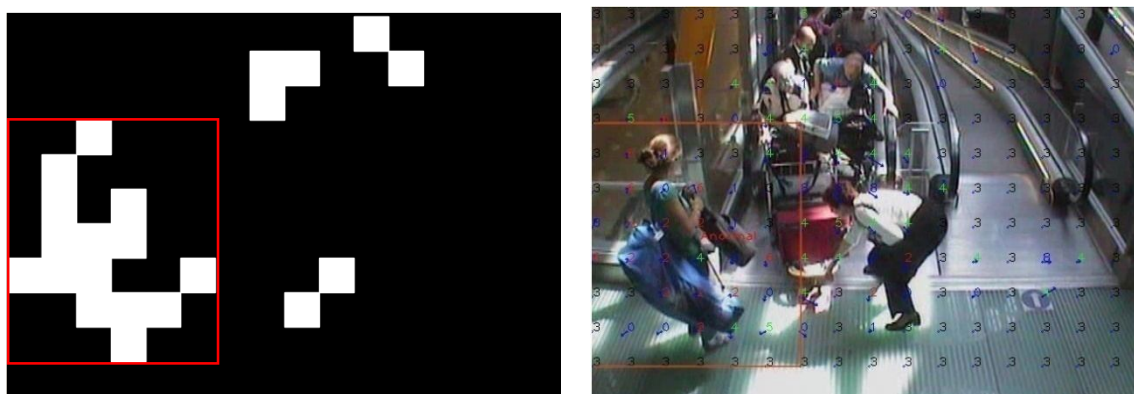


Figure 2.12 : Détection des blocs où le MCF est égale à 2, 6 et 10 dans l'image

Si le nombre de blocs contigus présentant une valeur anormale pour MCF, durant un nombre significatif de frames, un évènement anormal est signalé.

Le nombre d'occurrence des valeurs anormales du MCF (2, 6, 10 et 8 dans le cas d'un événement de bousculade) augmente en cours du temps. La figure 2.11 montre un graphe qui représente le nombre d'occurrence de la valeur MCF=2.

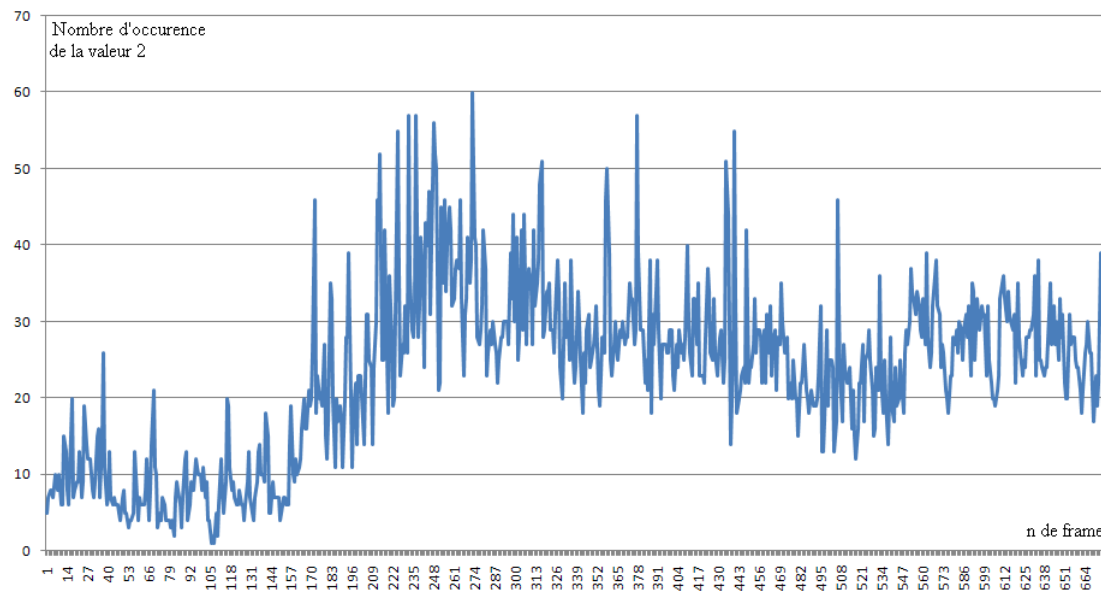


Figure2.13: Histogramme correspondant aux valeurs 2,6 et 10 du facteur MCF

Le principe de base est d'utiliser la matrice MCF de chaque image. Après avoir détecté les blobs (régions constitués d'un ensemble de blocs qui portent la même valeur ou la même sémantique) qui correspondent à une discontinuité, nous vérifions, pour chaque blob, s'il existe durant un nombre d'image significatif sans que sa taille ne diminue. Si c'est le cas, alors l'événement anormal est détecté.

5. Conclusion

Dans ce chapitre, nous avons présenté les aspects théoriques de l'approche proposée pour la détection d'évènements anormaux. Un exemple de tels évènements : la bousculade, la fuite...etc.

Nous avons proposé une décomposition fonctionnelle des traitements du système proposé constituée de trois niveaux : bas, intermédiaire et haut.

Quelques primitives ont été proposées pour la détection d'évènement anormal (bousculade) en se basant sur les primitives du niveau intermédiaire.

Dans le chapitre suivant, nous présenterons les résultats obtenus en appliquant notre système sur des vidéos de scènes réelles.

Chapitre 3

Résultats Expérimentaux

Dans ce chapitre, nous présentons les résultats obtenus suite à l'application des algorithmes implémentés sur des scènes de vidéo réelles. En raison du manque de données concernant ce type d'application, nous avons fait des tests sur des vidéos du laboratoire LIFL (université Lille, France) prises par des caméras de surveillance des escalateurs dans un aéroport. Nous avons utilisé pour une scène donnée, des vidéos ne contenant pas des événements anormaux pour pouvoir calculer les primitives de référence. Le second type de vidéos contient uniquement l'évènement anormal de bousculade.

1. Données expérimentales

Généralement, dans les vidéos traitées, nous avons deux escalateurs, correspondant à deux chemins dans deux directions opposées. L'ensemble des données est divisé en deux catégories : les vidéos de situations normales et de situation anormales. Les situations normales correspondent au mouvement de foule sans événement de bousculade dans l'escalateur. Les situations anormales correspondent aux vidéos qui contiennent un événement de bousculade dans l'escalateur.

2. Application des traitements

Dans la phase d'apprentissage, la taille de l'image dans la vidéo est de 640x480 pixels et chaque vidéo dans les deux catégories contient plus de 400 frames.

Pour la détection et le suivi de points d'intérêts, nous avons extrait environ 1500 points dans chaque frame (incluons les points statiques et ceux dus au bruit). Chaque frame de la vidéo est décomposée en un nombre de blocs de taille 40x40. Chacun de ces blocs contient un ensemble de vecteurs de mouvement (flot optique).

Après cette étapes, nous avons calculé pour chaque bloc les vecteurs de référence (au plus trois direction différentes). Pour la classification nous avons choisi le nombre de clusters égal à trois pour ce type de vidéos. Dans d'autres types de scènes le nombre de clusters change comme dans le cas d'un hall. Pour le fusionnement des vecteurs moyens obtenus, nous avons choisi un angle égal à 120 degrés pour séparer deux directions différentes.

Nous avons fait ce travail pour trouver les vecteurs de référence de mouvement pour chaque bloc de l'image.

Nous avons ensuite pris d'autres vidéos de la même scène et nous avons calculé pour chaque frame les vecteurs de mouvements moyens.

Nous avons ensuite évalué les différentes mesures : de direction et de magnitude pour chacun des blocs en utilisant les vecteurs de référence et les vecteurs moyens.

Pour signaler l'évènement anormal, il faut que le nombre de blocs ayant une valeur anormal de la mesure (MCF : Facteur de continuité de mouvement) soit important et reste constant pour une durée spécifique de temps. Pour cela nous avons choisi une durée de 5 secondes. Chaque seconde dans les vidéos est constituée de 9 frames, donc, l'alarme se déclenche après 45 frames.

3. Les résultats de détection d'évènement anormal

Le résultat expérimental montre que les événements anormaux sont détectés avec succès. Les figures 3.1, 3.3 et 3.5 montrent quelques frames de la vidéo contenant un évènement anormal. Les figures 3.2, 3.4 et 3.6 représentent les valeurs anormales de la mesure MCF associées aux blocs de plusieurs frames qui correspondent à l'évènement de bousculade détecté. Les blocs où il n'y a pas de mouvement sont indiqués par la valeur 3. Nous pouvons voir dans les frames qu'il y a plusieurs blocs contigus avec des valeurs égales à 2, 6, 8, et 10 qui sont synonymes d'un évènement anormal.

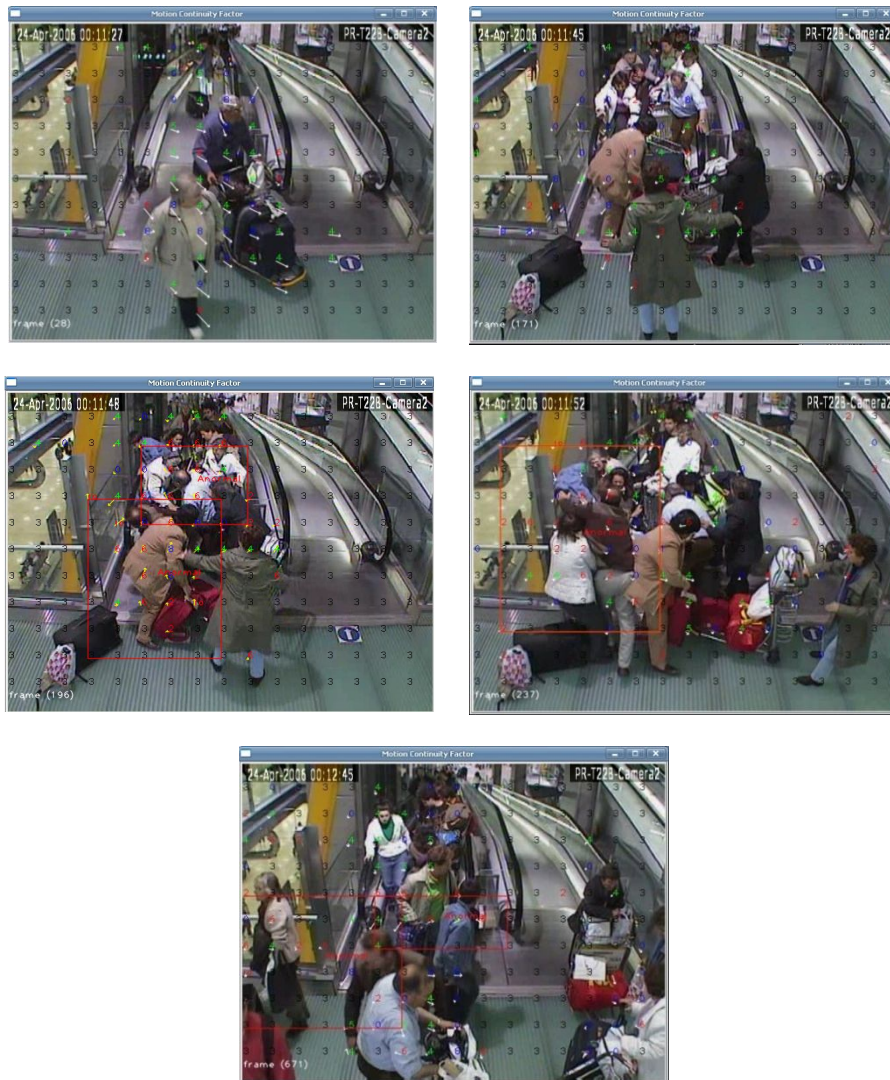


Figure 3.1: Evènement anormal qui commence à partir du frame 196.

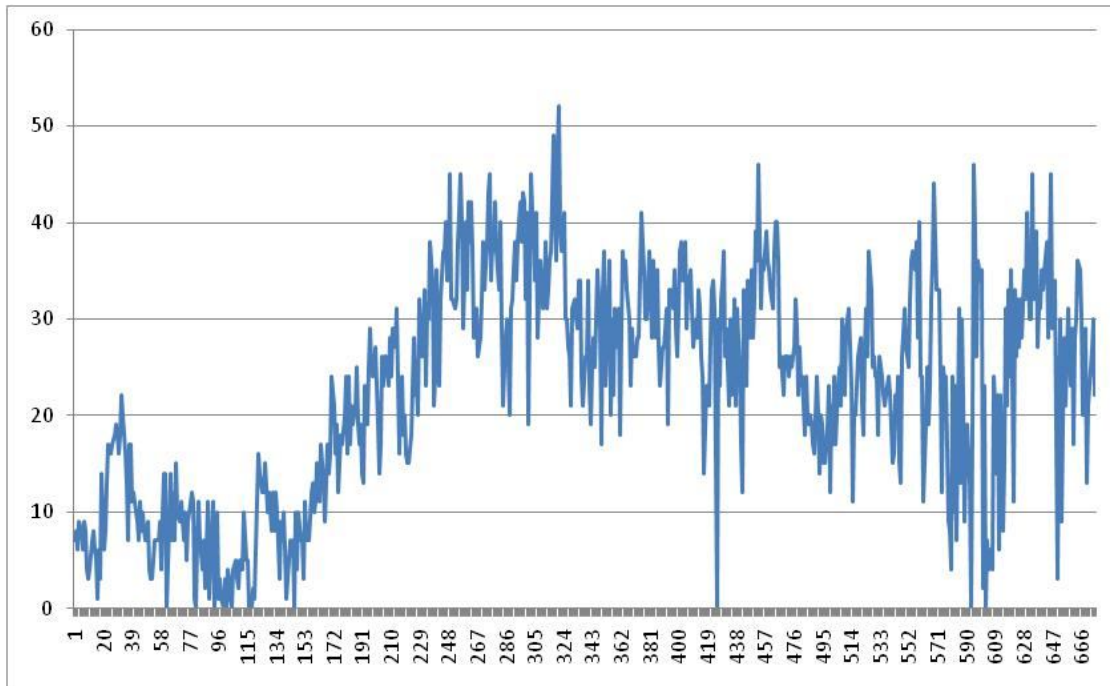


Figure 3.2 : Nombre d'occurrences des valeurs anormales de MCF pour les différents frames de la vidéo.

A partir du graphe de la figure 3.2, l'événement anormal illustré par la figure 3.1 est détecté à partir du frame 157.





Figure 3.3: Ev nement anormal qui commence   partir du frame 157.

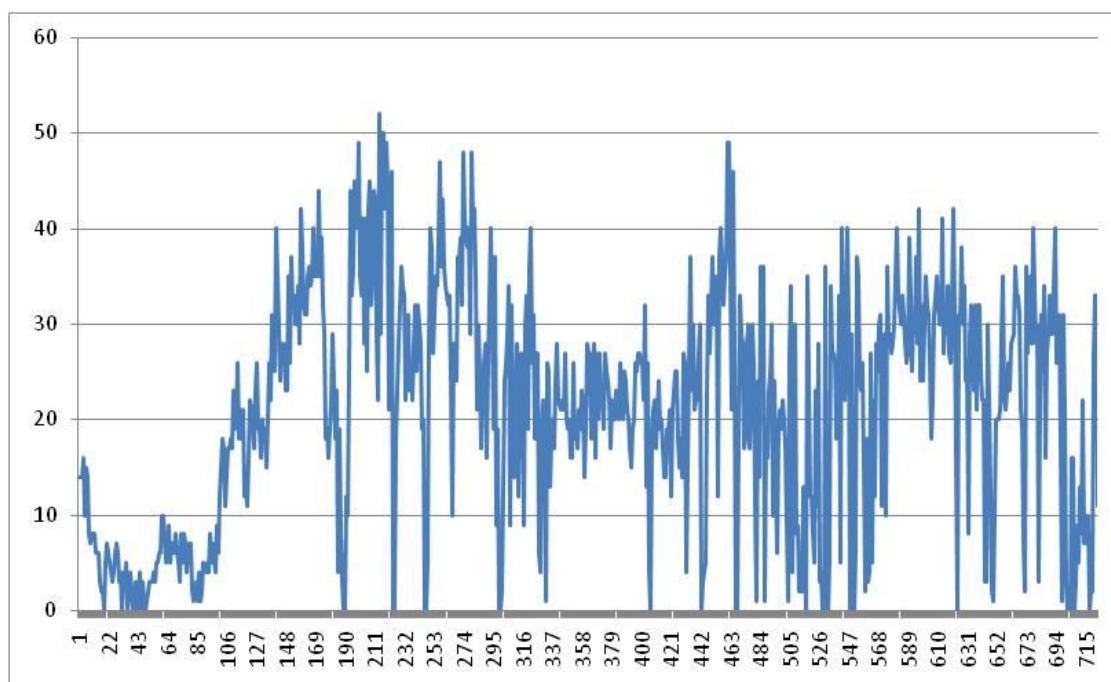


Figure 3.4 : Nombre d'occurrences des valeurs anormales de MCF pour les diff rents frames de la vid o



Figure 3.5: Evènement anormal qui commence à partir du frame 133.

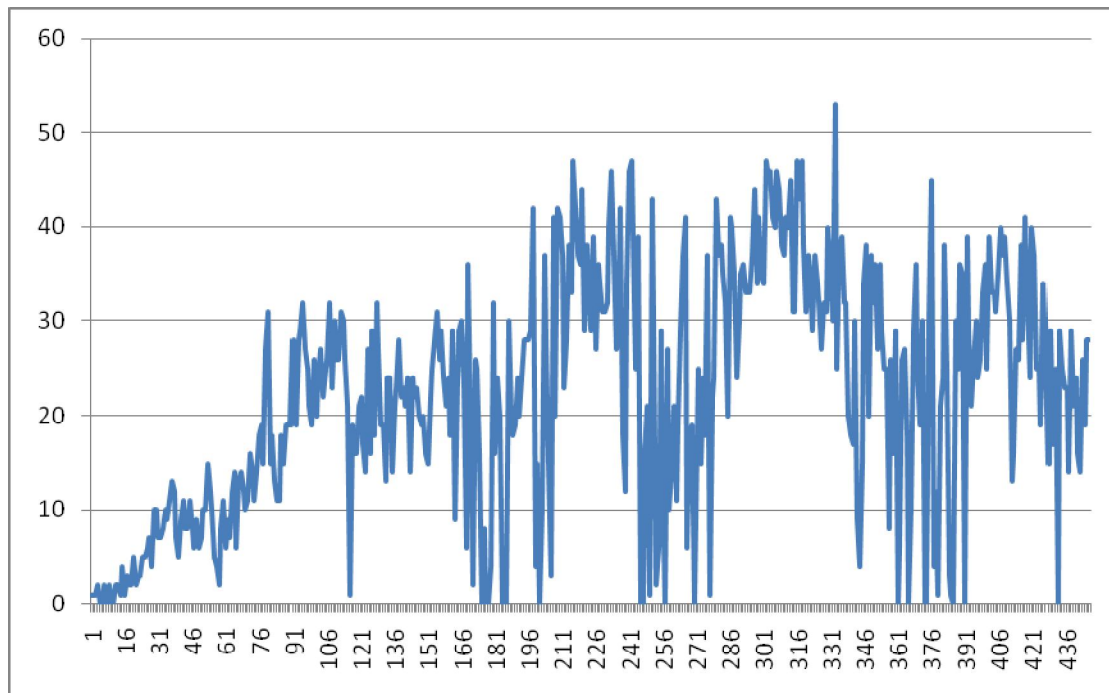


Figure 3.6 : Nombre des valeurs anormales de MCF pour les différents frames de la vidéo

Les événements anormaux détectés par le système sont comparés avec des événements anormaux annotés manuellement, qui montrent que les résultats sont satisfaisants. Notre approche n'est pas coûteuse en temps de calcul et espace mémoire.

La phase d'apprentissage est effectuée une fois pour une scène donnée. Les seuils doivent être choisis avec soin selon la scène.

4 Conclusion

Nous avons présenté dans ce chapitre les tests effectués sur des vidéos d'un escalateur. Nous avons commencé par le calcul des mesures normales pour les vidéos historiques de la scène. Nous avons ensuite, appliqué les algorithmes proposés pour les vidéos de scènes contenant des événements anormaux.

Concrètement, nous avons extrait pour chaque frame les points d'intérêt, calculé les vecteurs de mouvement et les vecteurs de mouvement moyens pour chacun des blocs de l'image.

Nous avons ensuite calculé les valeurs du facteur de continuité du mouvement en utilisant l'historique du mouvement.

Les résultats obtenus sur des vidéos de scènes réelles sont satisfaisants. Les mouvements anormaux contenus sont détectés.

Conclusion générale

Le développement de systèmes informatiques intelligents de vidéo surveillance est un thème d'actualité et de grand d'intérêt pour la communauté scientifique. Dans le cadre de ce mémoire, nous avons proposé une contribution au développement de tels systèmes. Concrètement, nous nous sommes intéressés à la détection d'évènements anormaux dans une scène à partir de la vidéo.

Au lieu de définir analytiquement un évènement anormal, qui est une tâche très complexe, nous avons opté pour l'utilisation d'un système qui travaille à partir d'une connaissance acquise par apprentissage réalisé sur des vidéos ne représentant pas les mouvements anormaux. Nous avons ensuite travaillé sur d'autres vidéos actuelles. Nous avons comparé la connaissance extraite avec celles apprises. Dans le cas d'un écart entre les deux connaissances, un évènement anormal est signalé.

L'approche que nous avons proposée est articulée sur trois niveaux de traitement, le niveau bas, intermédiaire et haut.

Dans le niveau bas, un ensemble de caractéristiques telles que les points d'intérêt, le flot optique sont extraits. Une méthode de classification appelée K-means est utilisée pour classer les vecteurs de mouvement en une ou deux directions dominantes. Dans le niveau intermédiaire nous avons calculé la variation de la direction (sens) et de la magnitude (vitesse) pour distinguer la continuité du mouvement de sa discontinuité. Dans le haut niveau, nous avons introduit une mesure qui est la combinaison entre la variation de la direction et la magnitude. Cette mesure donne des valeurs caractérisant l'évènement anormal.

Après les tests effectués de notre méthode sur des vidéos d'un escalateur pris dans un aéroport, les résultats ont montré une détection des événements anormaux après une durée de temps de six secondes.

Comme perspectives, nous proposons le calcul de nouvelles primitives de niveau intermédiaire et de haut qui permettront de détecter d'autres types d'évènement : Dépôt d'objets, bagarre, vol, etc.

Annexe 1

Méthode Lukas et Kanade avec pyramide d'image

La méthode de flot optique Lukas et Kanade est une méthode à deux différentielles d'images, car elle a besoin de deux images pour fonctionner. C'est la méthode la plus populaire d'estimation de flot optique, elle a été développée par [Bruce D. Lucas](#) and [Takeo Kanade](#) [LUC81]. Elle introduit une nouvelle hypothèse qui nous autorise à avoir un système entièrement déterminé : donnant un pixel P, tous les pixels au voisinage de P sont supposés avoir la même vitesse que P.

Comme vu dans le chapitre 2, l'équation (Eq.2.7) de la contrainte du flot optique est à deux variables inconnues (V_x et V_y). où, V_x et V_y sont les composantes de vitesse du flot optique associées au pixel considéré. Les termes $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ et $\frac{\partial I}{\partial t}$ sont les dérivées de l'image en (x, y, t) qui peuvent être écrits comme I_x , I_y et I_t . La formule devient alors:

$$I_x V_x + I_y V_y = -I_t \quad (A.1)$$

L'équation A.1 peut être transformée en produit de matrices comme suit [GIU09]:

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = -I_t \quad (A.2)$$

La méthode LK suppose que le mouvement entre deux images est très petit et donc le déplacement d'un pixel entre les deux images est très faible. En modifiant l'équation (A.1), et en supposant (V_x, V_y) constant dans le voisinage (fenêtre carrée de m pixels) centré au pixel P, nous obtenons:

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ I_{x3} & I_{y3} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = - \begin{bmatrix} I_{t1} \\ I_{t2} \\ I_{t3} \\ \vdots \\ I_{tn} \end{bmatrix} \quad (A.3)$$

Où les I_{xi} , I_{yi} et I_{ti} représentent les pixels de l'image considérée. Donc, nous avons un plus grand nombre d'équations que de nombre d'inconnues, ceci permet de résoudre les équations.

En utilisant les notations suivantes :

$$A = \begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ I_{x3} & I_{y3} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \quad (A.4)$$

$$\vec{V} = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \quad (A.5)$$

$$b = - \begin{bmatrix} I_{t1} \\ I_{t2} \\ I_{t3} \\ \vdots \\ I_{tn} \end{bmatrix} \quad (A.6)$$

L'équation (A.2) devient:

$$A\vec{V} = b \quad (A.7)$$

Le but de la méthode de Lukas et Kanade (LK) est de minimiser $\|A\vec{V} - b\|^2$ en utilisant la méthode des moindres carrés. Par multiplication avec la transposée de la matrice A, le terme $A\vec{V} - b$ peut être écrit comme suit :

$$A^T A \vec{V} = A^T b \quad (A.8)$$

A partir de l'équation A.8, nous obtenons les composantes du mouvement V_x et V_y . Nous écrivons la dernière équation avec plus de détails et nous obtenons :

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \vec{V} = \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} b \quad (A.9)$$

Où la matrice $A^T A = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix}$ est inversible et les valeurs propres sont non nuls.

La solution de l'équation A.8 est donc

$$\vec{V} = (A^T A)^{-1} A^T b \quad (A.10)$$

Le terme $(A^T A)^{-1} A^T$ est la pseudo-inverse de la matrice A et peut être calculée en utilisant les valeurs d'intensité I_x et I_y de tous les pixels de chaque image [GIU09] [RIS08]. Et $\vec{V} = [V_x, V_y]$, avec V_x et V_y sont les composantes du mouvement que nous recherchons.

Références

- [ALI07] S. Ali and M. Shah. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. *Computer Vision and Pattern Recognition*, 2007. CVPR '07. IEEE Conference on, pages 1{6, 17-22 June 2007.
- [AND08] E. L. Andrade, R. B. Fisher and S. Blunsden. Detection of emergency events in crowded scenes, *Proc. Of IEE Int. Symp. on Imaging for Crime Detection and Prevention*, pp. 528-533, 2006.
- [ARA01] J. Aranda J. Verges-Llahi and A. Sanfeliu, "*Object tracking system using colour histograms*", 9th Spanish Sym. Pattern Recog. Image Anal., Castellon, May 2001.
- [BEA95] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Computing Surveys* 27 (1995): 433–466.
- [BOI07] O. Boiman and M. Irani. Detecting irregularities in images and in video. *International Journal of Computer Vision*, 74(1):17{31, 2007.
- [BOU00] J.-Y. Bouguet. Pyramidal implementation of the Lucas-Kanade feature tracker. Technical report, Intel Corp., Microprocessor Research Labs, 2000.
- [CHE04] S.-C. S. Cheung and C. Kamath. *Robust techniques for background subtraction in urban traffic video*. In *Proc. SPIE Visual Communications and Image Processing*, pages 881 892, 2004.
- [CUP04] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat. Video understanding for metro surveillance. *Networking, Sensing and Control*, 2004 IEEE International Conference, vol. 1:186{191, 2004.
- [DAV03] A. Y. N. David M. Blei and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

- [DAV97] J. W. Davis and A. F. Bobick. The representation and recognition of action using temporal templates. M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 402.
- [DEV06] F. Devernay, D. Mateus et M. Guilbert, Multi-Camera Scene Flow by Tracking 3-D Points and Surfels, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, IEEE, New York, 2006.
- [DUD01] R. Duda, P. Hart, and D. Stork, “Pattern Classification,” by John Wiley & Sons, Inc. pp. 550-556, 2001.
- [GAR08] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O’Reilly Media, 1 edition, 10 2008.
- [GIU09] Giuseppe Catalano Alessio Gallace Bomi Kim Sergio Pedro Francesco Santoro Bomi Kim. Optical Flow.2009.
- [HAL05] D. Hall, J. Nascimento, P. C. R. E. Andrade, P. Moreno, S. P. T. List, R. Emonent, R. B. Fisher, J. Santos-Victor, and J. L. Crowley. *Comparision of target detection algorithms using adaptive background models*. In *Proc. IEEE Workshop on VS-PETS*, pages 113–120, 2005.
- [HAN07] Chin-Chuan Han, Cheng-Yi Lin, Gang Feng Ho and Kuo-Chin Fan, “Abnormal event detection using Trajectory Features”.
- [HEL95] D.Helbing and P.Molnar. Social force model for pedestrian dynamics. *Phys. Rev E*, 51(5) :4282, 1995.
- [HOR81] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence* 17 (1981): 185–203.
- [HUA95] Y. Huang and X. H. Zhuang, “Motion-partitioned adaptive block matching for video compression,” *International Conference on Image Processing* (vol. 1, p. 554), 1995.
- [JEA02] Jean-Yves Bouguet, “Pyramidal Implementation of the Lucas Kanade Feature Tracker – Description of the algorithm”, Intel Corporation – Microprocessor Research Labs, 2002.
- [JIA07] Fan Jiang, Ying Wu, Aggelos K. Katsaggelos, “Abnormal event detection from surveillance video by dynamic hierarchical clustering”, *ICIP* 2007.
- [JON09] Jonathon Shlens_Center for Neural Science, New York University New

- York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk
Insitute for Biological Studies
La Jolla, CA 92037. Avril 2009.
- [KAN02] Kanungo, T.; Mount, D. M.; Netanyahu, N. S.; Piatko, C. D.; Silverman, R.; Wu, A. Y. (2002). "An efficient k-means clustering algorithm: Analysis and implementation". *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**: 881–892.
- [LAW89] Lawrence R. Rabiner (February 1989). "A tutorial on Hidden Markov Models and selected applications in speech recognition". *Proceedings of the IEEE* **77** (2): 257–286.
- [LOW04] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* 60 (2004): 91–110.
- [LUC81] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 1981 DARPA Imaging Understanding Workshop* (pp. 121–130), 1981.
- [MOR79] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* 60 (2004): 91–110.
- [MOS1] Mosig. « Détection de points d'intérêts. »
<http://perception.inrialpes.fr/people/Boyer/Teaching/Mosig/Pinterets.pdf>;
- [MOS2] Mosig. « Détection des contours. »
<http://perception.inrialpes.fr/people/Boyer/Teaching/Mosig/contours.pdf>.
- [OLI00] N. M. Oliver, B. Rosario, and A. Pentland. A Bayesian computer vision system for modeling human interactions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):831–843, 2000.
- [OWN00] J.Owens and A. Hunter, "Application of the self-organizing map to trajectory classification" in Proc. Of the Third IEEE International Workshop on Visual Surveillance, 2000.
- [PAR00] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Mars 2000.
- [PIC04] M. Piccardi. Background subtraction technique: review. In Proc. IEEE Intern. Conf. on Systems, Man and Cybernetics, volume 4, pages 3099-3104, 2004.

- [PIE05] Pierre Gabriel, Jean-Bernard Hayet, Justus Piater et Jacques Verly, "*Utilisation des Points d'Intérêts Couleurs pour le Suivi d'Objets*", Actes de congrès ORASIS, France, 2005.
- [RAM09] R. Mehran, A. Oyama and M. Shah. Abnormal Crowd Behavior Detection using Social Force Model. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, Florida, 2009.
- [RIS08] Rishabh Malhotra: "Detection and Segmentation of Moving Objects in Video Using Optical Vector Flow Estimation", M.Sc. Thesis, University of Saskatchewan, Department of Electrical and Computer Engineering, July 2008.
- [SHI94] J. Shi and C. Tomasi, "*Good features to track*", in CVPR'94, pages 593–600, Seattle, Washington, June 1994.
- [SHI94] J. Shi and C. Tomasi, "*Good features to track*," *9th IEEE Conference on Computer Vision and Pattern Recognition*, June 1994.
- [STA99] C. Stauffer and W. Grimson. *Adaptive background mixture models for real-time tracking*. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume II, pages 246–252, 1999.
- [THI05] Thi-Thanh-Hai Tran et Augustin Lux, "*Suivi à base des crêtes d'objets mobiles dans une séquence vidéo*", in Intl. Conf. RIVF'04, Hanoi, Vietnam, February 2005.
- [VEL01] B.P.L LO and S.A. Velastin, "*Automatic congestion detection system for underground platforms*" *Proc. ISIMP 2001*, pp 158-161, May 2001.
- [WRE97] C. Wern, A. Azarbayejani, T. Darrell, and A.P. Pentland, "*Pfinder: real-time tracking of the human body*", *IEEE Trans. On Pattern Anal. And Machine Intell*, vol. 19, no 7, pp. 780-785, 1997.
- [XIN05] Xingzhi Luo and Suchendra M. Bhandarkar. "*Tracking of Multiple Objects Using Optical Flow Based Multiscale Elastic Matching*", in international conference of computer vision, (iccv2005), Beijing and China ,21 October 2005.

