

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
Université Ahmed Draia - Adrar
Faculté des Sciences et de la Technologie
Département des Mathématiques et Informatique



Mémoire de fin d'étude, en vue de l'obtention du diplôme de master en
informatique

Option: Réseaux et Système Intelligents

Thème

*Proposition d'une approche intelligente pour la
reconnaissance d'actions humaines à partir d'image de
vidéosurveillance*

Préparé par

M^{elle} Nadia BOUTADARA et M^{elle} Fatima Zahra BOUAZZA

Encadré par

Mr. KOHILI Mohammed

Année Universitaire 2016/2017

Remerciement

Avant de commencer la présentation de ce travail, Nous tenons à remercier en premier notre DIEU tout puissant de nous avoir donné le courage, la patience, la foi et la volonté pour pouvoir continuer et réaliser ce travail.

Nous tenons tous d'abord à remercier nos parents pour leurs soutiens et leurs encouragements durant nos années d'études.

Un remerciement spécial pour notre grand et respectueux professeur M.KOHLI Mohammed, d'avoir accepté de nous 'encadrer pour notre projet de fin d'études, ainsi que pour son soutien, ses remarques pertinentes et son encouragement.

À tous nos enseignants depuis primaire jusqu'à notre cursus universitaire.

Nous profitons de l'occasion pour remercier toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet de fin d'études.

BOUTADARA Nadia

&

BOUAZZA Fatima Zahra

Dédicace

Je dédie ce modeste travail à :

A les lumières de mes jours, les sources
de mes efforts, Les flammes de
Mon cœur, ma vie et mon bonheur ; *ma mère*
et *ma grand-mère* que j'adore.

A l'homme de ma vie, mon exemple éternel, mon soutien
moral et source de joie et de bonheur, celui qui s'est
toujours sacrifié pour me voir réussir *mon père*.

Tous mes chers frères et mes sœurs : Ahmad, Abdelkader,
Abdeassalam, Fatima, Zenab.

A mes tantes et ses fils et mes oncles: Ibrahim, Mohammad,
Abdeassalam.

mes chères amies :Rekia, Nacira, Meriem, Sayida, Hanan,
Malika, Saida, Khadidja, A_Souaad, Fatima, Zinab ...

A ma chère amie, ma binôme *Fatima Zahra*.

A tous ma famille **BOUTADARA**.

Tous mes enseignants depuis primaire
jusqu'à mon cursus universitaire.

Toute la promotion 2016/2017.

Nadia



Dédicace

Que ce travail témoigne de mes respects :

*Au printemps compassion inépuisable; à la mémoire
de ma chère grand-mère ♥'Ma_Fatma.*

*A la plus belle créature que Dieu a créée sur terre,,
A cet source de tendresse, de patience et de générosité,,*

A ma mère ♥

A mon père ♥:

*Grâce à leur tendre encouragement et leur grand
sacrifice, il a pu créer le climat affectueux et propice
à la poursuite de mes études.*

*A tous mes chers frères♥: Mohammed, Ibrahim,
Hamza, Mustafa et le petit Nourelddine.*

*Je prie le bon Dieu de les bénir, de veiller sur eux,
en espérant qu'ils seront toujours fiers de moi.*

A mes Tantes et mes oncles;

A mes chères amies et sœurs ...

A tous mes amis et mes collègues.

♥ *A ma belle amie, ma binôme Nadia♥.*

A tous ceux qui, par un mot,

*m'ont donné la force ♥
de continuer*

*Au temps qui passe,
à ma vie ♥...*

♥ *Fatima Zahra* ♥

Résumé: Dans cette mémoire, nous proposons une méthode de reconnaissance d'activités, fondée sur un modèle hybride Séparateur à Vaste Marge-Chaîne de Markov Cachée (SVM-MMCs) qui modélise explicitement l'aspect séquentiel des activités tout en exploitant des paramètres géométriques qu'on a les proposées, le jeu des expériences était extrait d'une base de données vidéo **KTH**. Nous avons montré, en effet, que l'utilisation de cette hybridation permet d'améliorer la performance du système de reconnaissance.

Mots clés : Reconnaissance d'activités ; vidéo surveillance, Classification de données séquentielles ; Séparateurs à vaste marge ; Chaines de markov cachés, vidéo, reconnaissance.

Abstract: In this memory, we propose a method of activity recognition, based on a hybrid model of Hidden Markov-Chain Separator (SVM-HMMs) which explicitly models the sequential aspect of the activities while exploiting geometric parameters, that we have proposed, the game of the experiments was playing on a video database KTH. We have shown that the use of this hybridization makes it possible to improve the performance of the recognition system.

Keywords: Recognition of activities; Video surveillance, sequential data classification; Large margin separators; Hidden markov Model, video, recognition.

ملخص: في هذه الذاكرة، نقترح طريقة للتعرف على الأنشطة، استناداً إلى نموذج مركب بين فاصل الهوامش الواسعة وسلسلة ماركوف المخفية (SVM-المركبه) التي تعتمد على مجموعة من الخصائص الهندسية، وهي الطريقة التي اقترحناها، تم استخدام قاعدة بيانات الفيديو KTH لإجراء مجموعة من التجارب. لقد أظهرنا، في الواقع، أن استخدام هذا التركيب يحسن أداء نظام التعرف.

الكلمات المفتاحية: التعرف على الأنشطة؛ المراقبة بالفيديو، متابعة تصنيف البيانات؛ مع فواصل هامش واسع. سلاسل ماركوف المخفية، والفيديو، والتعرف.

Table de matière:

Résumé.....	I
Table de matière.....	II
Liste des figures.....	V
Liste des tableaux.....	VI
Introduction générale.....	1
Chapitre 01: Généralités sur la vidéo surveillance	
1.1. Introduction.....	2
1.2. L'image numérique.....	2
1.2.1. Définition de l'image	2
1.2.2. Représentation d'une image	2
1.2.3. Les types d'image numérique.....	2
1.2.4. Les caractéristiques d'image	3
1.3. Vidéo.....	4
1.3.1. Définition	4
1.3.2. Frame.....	5
1.3.3. Objet.....	5
1.3.4. Types de vidéo	5
a. La vidéo analogique.....	5
b. La vidéo numérique.....	5
1.3.5. Les paramètres clés d'une vidéo	5
1.4. La vidéosurveillance.....	6
1.4.2. Principe d'un système vidéo surveillance	6
1.4.3. Objectifs	7
1.4.4. Les problèmes liés à la vidéosurveillance	7
1.4.5. Les types de vidéosurveillance.....	7
a- Vidéosurveillance analogique.....	7
b- Vidéosurveillance IP.....	8
c- Passer de l'analogique au numérique : pas difficile.....	9
1.4.6. Domaines d'application de vidéosurveillance.....	11
1.5. Conclusion.....	12

Chapitre 02 : Méthodes de classification

2.1.	Introduction.....	13
2.2.	Les Machines à Vecteurs de Support (SVM)	13
2.2.1.	Introduction	13
2.2.2.	Historique	13
2.2.3.	Définition	14
2.2.4.	Principe de fonctionnement.....	15
2.2.5.	Notions de base	16
	a- Hyperplan.....	16
	b- Vecteur supports	16
	c- Marge.....	16
2.2.6.	Linéarité et non-linéarité	16
2.2.7.	Cas non linéaire.....	17
2.2.8.	Noyaux SVM.....	18
2.2.9.	Avantage et Inconvénients	18
2.2.10.	Domaines d'application d' SVM.....	18
2.3.	Chaînes de Markov cachées.....	19
2.3.1.	Introduction	19
2.3.2.	Définition	19
2.3.3.	Processus stochastique	20
2.3.4.	Présentation et applications des modèles de Markov cachés	20
2.3.5.	Les algorithmes de MMCs	19
	a- Algorithme forward-backward.....	19
	b- Algorithme de Viterbi.....	20
	c- Entraînement de Baum-Welch.....	20
2.3.6.	Apprentissage d'un MMCs	23
2.3.7.	Différence types de MMCs	24
2.3.8.	Les avantages et Les inconvénients :	25
	a- Les avantages	25
	b- Les inconvénients.....	26
2.3.9.	Domaine d'application	26
2.4.	Conclusion.....	26

Chapitre 03 : Implémentation et expérience

Introduction.....	27
3.2.Environment de programmation utilisée	28
3.3.La base de données.....	28
3.4.Architecture du système proposé	30
a- Phase d'apprentissage SVM :.....	30
b- Phase d'apprentissage hybridation SVM-MMC :	31
c- Phase de reconnaissance :	32
3.5.Explication détaillée des états du système	33
3.5.1.Lecture de séquences vidéo.....	33
3.5.2.Phase de discrétisation de vidéo.....	34
3.5.3.Détection d'objet.....	34
3.5.4.Phase de prétraitement	35
3.5.5.Mesure de similarité.....	36
3.5.6.Modèle SVM.....	36
3.5.7.Graphe de décision	36
3.5.8.Construction des modèles MMCs	38
3.5.9.Décision.....	39
3.6. Analyse des résultats	39
3.6.1.La décision sur la vidéo globale	40
a. La méthode de classification avec SVM non globale.....	40
b. La méthode de classification avec SVM linéaire	41
c. La méthode de classification avec SVM - MMCs.....	42
3.6.2.La décision sur les frames	43
a- SVM non linéaire	43
b- SVM linéaire	44
a- SVM – MMCs	44
3.7.Etude comparatif	45
3.8.Conclusion.....	46
Conclusion générale	47
Référence bibliographique	48

Liste des figures

Figure 1.1: Image binaire

Figure 1.2: Image en niveaux de gris (Nuance de 256 gris Exemple de photo possible en 8bpp)

Figure 1.3: Image en couleurs

Figure 1.4 : Pixel dans une image numérique

Figure 1.5 : Voisinage à 4.

Figure 1.6 : Représentation d'un flux de vidéo.

Figure 1.7 : Schéma de principe de vidéosurveillance.

Figure 2.1 : Frontière de décision linéaire d'un classifieur SVM.

Figure 2.2 : Cas linéairement et non linéairement séparable.

Figure 2.3 : Hyperplan dans le cas des données non linéairement séparable.

Figure 2.4 : Exemple de Modèle de Markov caché pour l'action «walk»[24].

Figure 2.5 : modèle de chaîne de Markov ergodique

Figure 2.6: modèle de Bakis.

Figure 3.1: Illustration de données KTH. Echantillon pour les six classes d'actions (colonne)enregistrées sous différents scénarios (par rangée).

Figure 3.2:Structure de phase d'apprentissage SVM.

Figure 3.3:Structure de phase d'apprentissage pour le système hybride SVM-MMC.

Figure 3.4:Structure de phase de reconnaissance pour SVM et système hybride SVM-MMC.

Figure 3.5 : Représentation de lecture de chaque vidéo.

Figure 3.6 : Représentation d'extraction des images.

Figure 3.7 : Manière de nommage des images.

Figure 3.8 : Méthode de détection d'objet.

Figure 3.9 : Méthode d'élimination d'ombre.

Figure 3.10 : Explication des caractéristiques.

Figure 3.11 : Structure de fonctionnement d'un SVM de type un contre un.

Figure 3.12 : Structure de modèle MMCs.

Figure 3.13:Schéma représentela méthode de prendre la décision.

Liste des tableaux

Tableau 3.1: Matrice de confusion de SVM non linéaire kernel RBF avec $g_a= 0.5$.

Tableau 3.2: Matrice de confusion de SVM non linéaire kernel RBF avec $g_a= 1.5$.

Tableau 3.3: Matrice de confusion avec SVM linéaire.

Tableau 3.4: Matrice de confusion avec hybridation SVM-MMCs

Tableau 3.5: Matrice de confusion de SVM non linéaire kernel RBF avec $g_a= 0.5$, nombre des images 10

Tableau 3.6 : Matrice de confusion avec SVM linéaire avec frames, nombre d'images=10

Tableau 3.7: Matrice de confusion avec hybridation SVM-MMCs, en utilisant des frames de dix images

Tableau 3.8: Comparaison des performances sur la base KTH.

A large, light orange decorative shape with a wavy, irregular border, resembling a stylized letter 'C' or a shield-like form. It is centered on the page and contains the title text.

Introduction générale

Introduction générale

La reconnaissance automatique des actions humaines est un sujet particulièrement complexe dans le domaine de la vision par ordinateur, elle consiste en la classification automatique des actions ou des activités réalisées par un individu dans une séquence vidéo. Un grand nombre d'applications de reconnaissance des actions humaines à partir des vidéos peuvent être trouvées : la vidéo surveillance, l'imagerie médicale, la robotique, l'interaction homme machine, l'analyse de séquences sportives...etc.

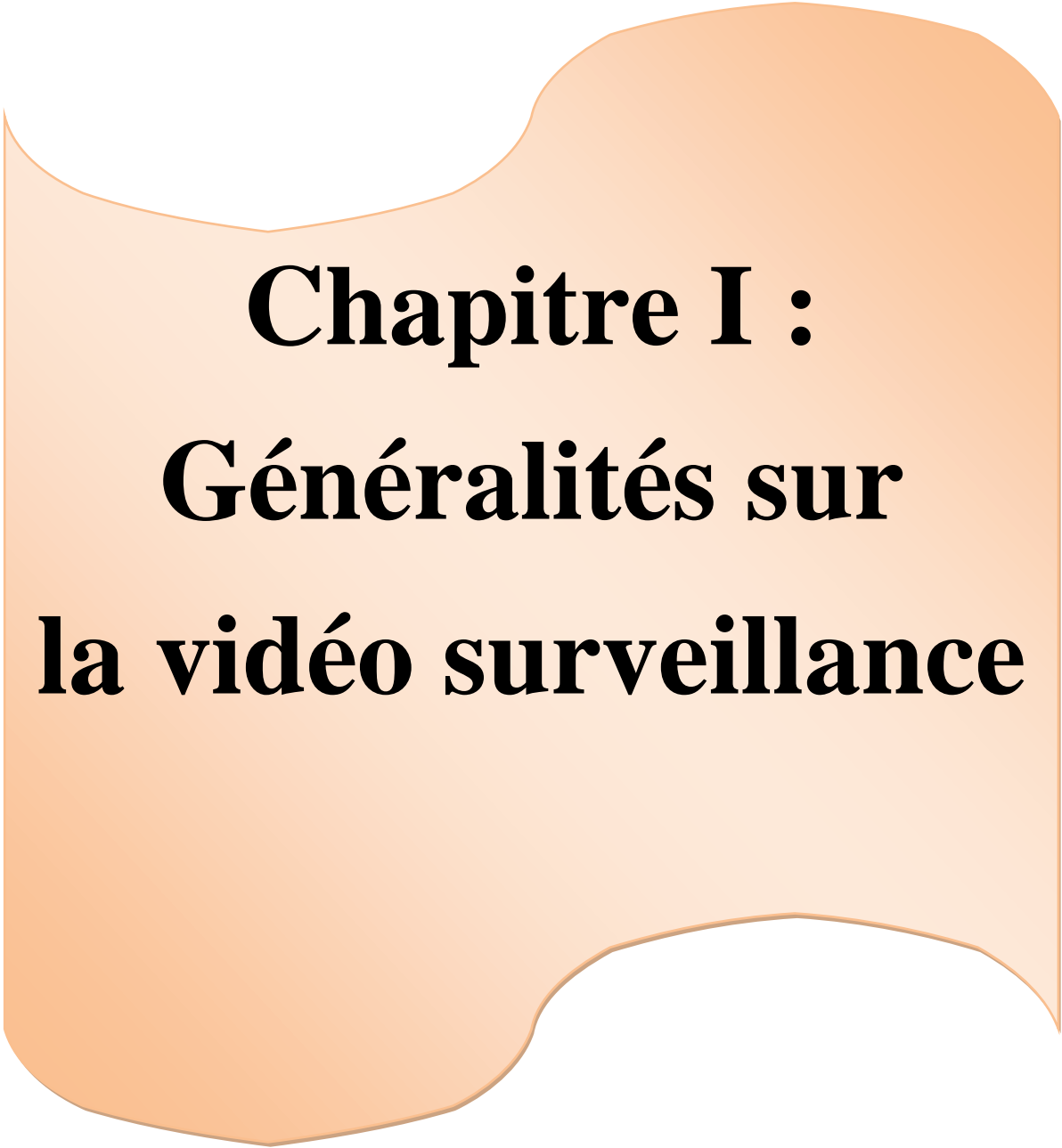
En effet, les zones de mouvement d'une séquence d'images correspondent souvent à des événements sur lesquels un système de vision doit se focaliser.

L'objectif d'un système de reconnaissance d'actions est de reconnaître des actions simples de la vie courante dans une vidéo (ex : marcher, répondre au téléphone, sauter...) à partir de vidéos de référence. Ces actions répondent à des modèles de mouvements simples effectués par une seule et même personne durant un laps de temps court.

La problématique de la reconnaissance d'action suscite l'intérêt depuis les premières bases vidéo et a énormément progressé ces dernières années. Les premiers travaux dans ce domaine consistaient à positionner, à la fois temporellement et spatialement, les actions dans la vidéo, alors que de nos jours on recherche à identifier ces actions. Les méthodes récentes d'identifications d'actions donnent de très bons résultats dans les bases vidéo d'actions en milieu contrôlé. Mais pour le moment, aucune méthode n'a fourni de résultats réellement satisfaisants dans les bases de vidéos réelles (films, vidéo surveillance, ...), Il est donc essentiel d'éprouver les nouvelles méthodes développées sur des bases de vidéos réelles.

Ce mémoire sera organisé de la manière suivante :

Dans le 1^{er} chapitre nous allons nous intéressons de définir une idée générale sur l'image numérique, puis on présente des généralités sur vidéosurveillance, ensuite la deuxième chapitre sera concerner en les modèles de Markov cachés et les machines à vecteurs de support (SVMs) ; leurs avantages et inconvénients. Après le troisième chapitre sera basé sur une présentation détaillée de notre application avec une les résultats de notre projet, les expériences et les discussions, En concluons par une conclusion générale.

A large, light orange decorative shape with a wavy, irregular border, resembling a stylized book cover or a shield. It is centered on a white background.

Chapitre I :

Généralités sur

la vidéo surveillance

1.1. Introduction

La vidéosurveillance est, le plus souvent, implantée dans le cadre d'un programme de prévention de la criminalité ou de renforcement de la sécurité publique.

Dans cette chapitre là ; nous allons présentons une vue globale sur l'image numérique et la vidéo surveillance.

1.2. L'image numérique

1.2.1. Définition de l'image

Une image est définie comme étant la représentation exacte ou analogique d'une scène par la peinture, la sculpture, le dessin, photographie, film...

Elle représente un ensemble structuré d'informations qui après affichage sur un support (écran, toile, papier, ...) ont une signification pour l'œil humain.

Si x et y sont les coordonnées spatiales d'un point de l'image, $I(x, y)$ est une fonction de l'intensité lumineuse et de la couleur [1].

Autrement dit; l'image numérique est l'image dont la surface est divisée en élément de taille fixe appelée cellules ou pixels. Une image peut être représentée sous quatre types.

1.2.2. Représentation d'une image

Elle peut s'effectuer selon diverses caractéristiques comme la couleur ou selon des coefficients tels les coefficients de Fourier ou d'ondelette. En informatique, le plus petit élément visuel représentable à l'écran se nomme pixel. Une image est donc représentée comme un tableau de pixels de différentes couleurs ou niveaux de gris [2].

1.2.3. Les types d'image numérique

- **Image binaire**

Une image binaire est une image $M \times N$ où chaque point peut prendre uniquement la valeur 0 ou 1. Les pixels sont noirs (0) ou blancs (1).

L'expression de la valeur d'image binaire $N_g = 2$ et la relation sur les niveaux de gris devient: $p(i,j) = 0$ ou $p(i,j) = 1$.

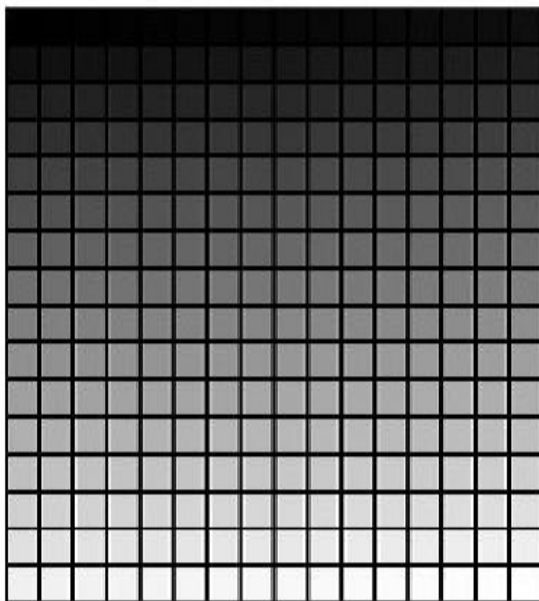
⇒ Chaque pixel peut donc avoir deux couleurs possibles : soit noir ou soit blanc

1	1	1	1	1	1	1	1	1	1
1	1	1	0	0	0	0	1	1	1
1	1	0	1	1	1	1	0	1	1
1	0	1	1	1	1	1	1	0	1
1	0	1	0	1	1	0	1	0	1
1	0	1	1	1	1	1	1	0	1
1	0	1	0	1	1	0	1	0	1
1	0	1	1	0	0	1	1	0	1
1	1	0	1	1	1	1	0	1	1
1	1	1	0	0	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1

Figure 1.1: Image binaire

- **Image en niveaux de gris**

Une image de niveaux de gris autorise un dégradé de gris entre le noir et le blanc. En général, on code le niveau de gris sur un octet (8 bits) soit 256 nuances de dégradé. Le niveau de gris est codé sur un bit (BinarydigIT).L'expression de la valeur du niveau de gris avec $N_g=256$ devient: $p(i,j) \in [0, 255]$ [3].



*Figure 1.2:*Image en niveaux de gris (Nuance de 256 gris Exemple de photo possible en 8bpp)

- **Image couleur**

Une image couleur est la composition de trois (ou plus) images en niveaux de gris sur trois (ou plus) composantes. On définit donc trois plans de niveaux de gris, un rouge, un vert et un bleu. La couleur finale est obtenue par synthèse additive de ces trois (ou plus) composantes.

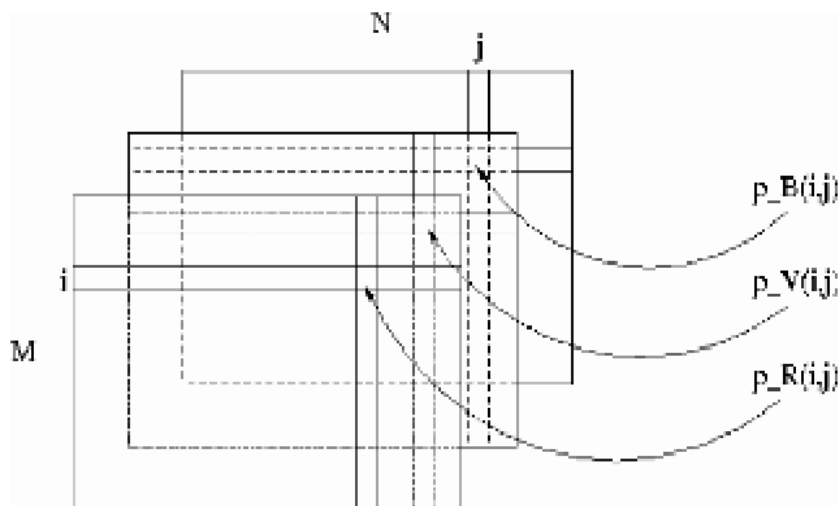


Figure 1.3: Image en couleurs

On a les relations sur les niveaux de gris: $p_R(i,j) \in [0, 255]$, $p_V(i,j) \in [0, 255]$, $p_B(i,j) \in [0, 255]$. On voit bien sur la figure qu'une image couleur est en fait l'association de trois plans de niveau de gris, chacun d'eux étant une couleur de base [3].

1.2.4. Les caractéristiques d'image [1]

- **Pixel :** Le mot pixel est l'abréviation du mot anglais « Picture element », il s'agit du plus petit élément constituant une image.

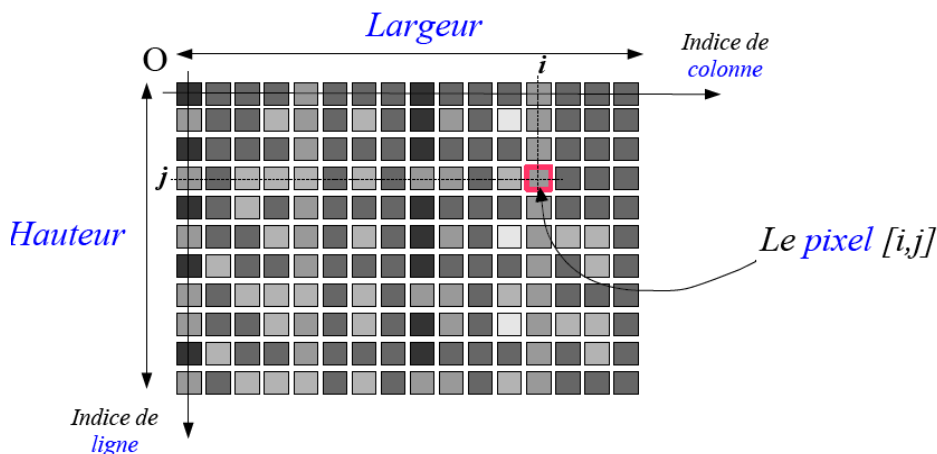


Figure 1.4 : pixel dans une image numérique

- **Voisinage:** C'est une notion fondamentale en traitement d'image, c'est l'ensemble des pixels qui se situent autour d'un pixel.

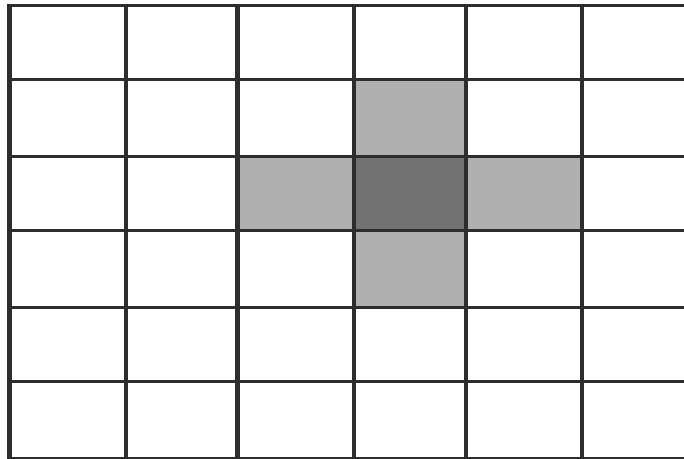


Figure 1.5 : Voisinage à 4

- **Dimension:** C'est la taille d'image, se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels).
- **Niveau de gris:** Le niveau de gris est la valeur de l'intensité lumineuse en un point. La couleur du pixel peut prendre des valeurs allant du noir au blanc en passant par un nombre fini de niveaux intermédiaires.
- **Luminance:** C'est le degré de la luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface.
- **Contraste:** Une image contrastée présente une bonne dynamique de la distribution des valeurs de gris sur tout l'intervalle des valeurs possible avec des blancs bien clairs et des noirs profonds.
- **Résolution:** Sur les ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure, aussi pour désigner le nombre total de pixels horizontaux et verticaux sur un moniteur.

1.3. Vidéo

1.3.1. Définition

Nous considérons un flux de vidéo comme une suite d'images 2D. La résolution de la vidéo, exprimée en nombre de pixels, définit la dimension de ces images.

La durée du temps entre deux images (Δt) est très petite parce que nous savons que la vitesse de film, en général, est de 24 à 60 images par seconde.

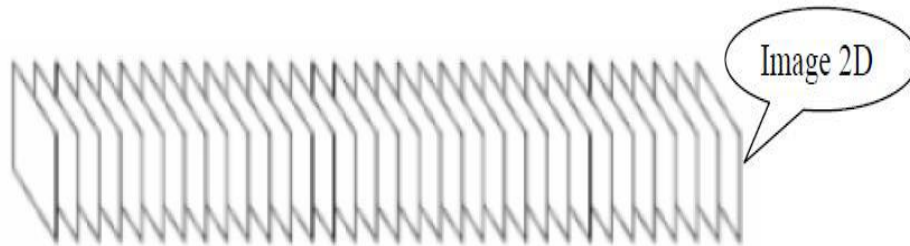


Figure 1.6 : Représentation d'un flux de vidéo

1.3.2. Frame

Image composants la vidéo, les photogrammes.

1.3.3. Objet

Les objets physiques sont les objets du monde réel qui apparaissent dans les scènes observées par les caméras.

Les objets physiques sont divisés en deux types : les objets de contexte et les objets mobiles.[4]

➤ **Les objets de contexte** : sont des objets physiques qui sont habituellement statiques (p. ex. les murs). Dans le cas où ils ne sont pas statiques, leurs mouvements peuvent être prédits par les informations contextuelles p.ex. les chaises, les portes sont des objets de contexte. [9]

➤ **Les objets mobiles** : sont des objets physiques qui peuvent être perçus dans les scènes par leurs mouvements. [9]

1.3.4. Types de vidéo [5]

Le signal vidéo est le signal qui permet de transporter une séquence d'images de la source à un dispositif d'affichage sous forme électrique. Selon la façon dont les signaux sont traités on peut distinguer les deux modes:

a. La vidéo analogique

Décrit le signal analogique comme un signal électrique dont l'intensité varie dans le temps de façon continue. La qualité du signal final dans ce mode est plus faible car le bruit rajouté au signal lors son traitement altère sa qualité.

b. La vidéo numérique

Est un signal qui porte une information représentée par une suite de valeurs minimales ou maximales correspondant respectivement au 0 et au 1. L'un des facteurs qui avantage le signal numérique par rapport au signal analogique est la facilité de distinguer l'information émise du bruit.

1.3.5. Les paramètres clés d'une vidéo

Le stockage et la diffusion d'une vidéo exigent un espace volumineux et un taux de transfert plus élevé. Le contrôle de qualité, et la taille d'une séquence vidéo est déterminé par deux paramètres clés,

le nombre d'images par seconde : le nombre d'images du système visuel humain exigé en général 25 ou 30 images par seconde.

La résolution : ce terme désigne que la quantité de l'information est limitée dans l'image. Autrement, c'est le nombre de pixels qui peuvent être affichés par un dispositif d'affichage.

Trouver le compromis entre ces paramètres et les limitations imposées par la technologie permet d'obtenir une qualité de vidéo optimale.

1.4. La vidéosurveillance**1.4.1. Définition**

La vidéo surveillance c'est un système de caméras permettant de surveiller à distance un espace privé ou public. Des images sont enregistrées avec ce système et sont par la suite visionnées et sauvegardées.[6]

Les systèmes de vidéosurveillance sont composés de différents types de matériel en fonction des besoins de son utilisateur (les caméras de surveillance, l'écran de vidéosurveillance, l'alimentation des caméras de vidéosurveillance, les enregistreurs de vidéosurveillance, Câble de vidéosurveillance ou liaison sans fil, .. etc). [7]

1.4.2. Principe d'un système vidéo surveillance

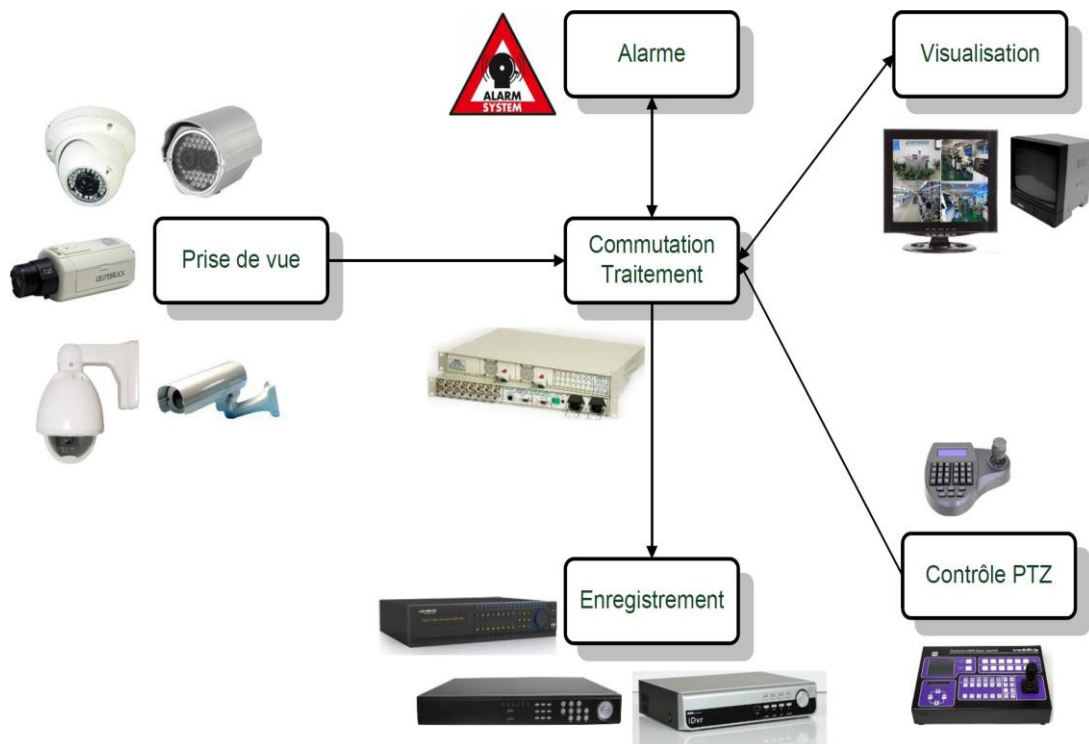


Figure 1.7: Schéma de principe de vidéosurveillance.

Ce schéma représente les différentes fonctions d'un système de vidéosurveillance. Celles-ci peuvent être assurées par un ou plusieurs appareils (Un DVR3 peut remplir les fonctions de commutation et d'enregistrement). Ce schéma est un schéma "par excès", certaines installations ne comportent pas autant de fonctions.[8]

1.4.3. Objectifs

L'objectif général d'un système de vidéosurveillance est de contribuer à la sécurité de biens et/ou de personnes.

Cette contribution peut se focaliser sur diverses composantes, sont les suivantes :

- Prévention de la criminalité
- Sécurité routière
- Sécurité industrielle
- Sûreté

1.4.4. Les problèmes liés à la vidéosurveillance[6]

Le plus important des problèmes est lié à l'atteinte de la vie privée puisque nous sommes constamment surveillés. Les familles ne se sentent pas à l'aise en sachant qu'une caméra les surveille en permanence. Cela pose une restriction des libertés.

L'autre problème est celui de la mise en place de caméra qui est très coûteuse puisque si l'on souhaite un système sophistiqué de vidéo surveillance cela représentera un investissement considérable et à cela s'ajoutera la maintenance du système qui devra être régulière. De plus, on ne peut pas surveiller toutes les rues d'une ville puisque pour le faire, il faudrait un nombre important de caméras. Enfin, cela enlève de l'emploi des vigiles dans les rues.

1.4.5. Les types de vidéosurveillance

Il existe deux types principaux de mode de vidéosurveillance : la vidéosurveillance IP et la vidéosurveillance analogique, plus traditionnelle.

a- Vidéosurveillance analogique

Le système de vidéosurveillance analogique est encore actuellement le plus utilisé dans les entreprises, commerces et habitations de particuliers. Il est équipé de caméras analogiques dont le seul rôle est de capturer les images et les envoyer à un enregistreur à durée limitée (type magnétoscope) via un signal analogique.

Ce système est composé d'une ou plusieurs caméras, d'un moniteur (ou téléviseur), d'un enregistreur et d'un câblage (le transfert d'images se fait via un câble dit coaxial).

Avantages de vidéosurveillance analogique

La position dominante de ce type de vidéosurveillance sur le marché apporte quelques avantages importants :

- son principal avantage est la qualité et la fluidité des images ainsi que sa facilité d'utilisation.
- une grande diversité de caméras analogiques : dimensions, formes, applications ;
- le prix des caméras analogiques : plus économique que celui des caméras IP.

Inconvénients de vidéosurveillance analogique

Par rapport à la vidéosurveillance en réseau, le système analogique est assez limité en termes de fonctionnalités :

- capacité de stockage limitée dans la durée (cassettes) ;
- format peu flexible des images ;
- pas d'accès en temps réel par l'Internet (sauf si les caméras sont connectées à un réseau) et l'Internet mobile (PDA, GSM...) ;
- évolutivité limitée : difficile d'ajouter des caméras supplémentaires dans le temps (longueur de câble, travaux...);
- pas de gestion à distance : installation, maintenance...

b- Vidéosurveillance IP

La vidéosurveillance IP (Internet Protocole) est venue compléter la vidéosurveillance analogique, qui équipe encore la majorité des installations.

Elle fonctionne avec les mêmes composants (caméras, moniteur, enregistreur, câbles) mais passe par un réseau informatique :

- les caméras IP sont installées sur un réseau IP (Intranet, Internet, LAN -réseau local-, CPL -courant porteur en ligne, c'est-à-dire les prises électriques - ou VPN...) et reliées à un serveur de vidéosurveillance central : elles capturent les images et les acheminent vers le réseau ;
- le serveur de vidéosurveillance est équipé d'un logiciel de vidéosurveillance : c'est lui le cœur du système, il récupère les images et les stocke sur disque dur.

Avantages de la vidéosurveillance IP

- Bénéficie de toutes les fonctionnalités d'Internet.
- Multiples grâce à la technologie réseau : la vidéosurveillance IP peut s'intégrer à d'autres technologies qui relèvent de l'IP (système de sécurité, visioconférence...)
- Peut s'installer sur un réseau informatique qui existe déjà (économie)
- Compatible avec tous types de câblage : IP, coaxial... ou sans fil (WiFi)
- Tout passe par un ordinateur : pas besoin de moniteur, d'enregistreur...
- Grande flexibilité d'installation des caméras : possibilité d'en ajouter facilement ou de les changer de place.

- Visualisation des images en temps réel
- Réglage possible des images (dimension, zoom...)
- Les caméras peuvent être commandées à distance (selon les modèles)
- Pas de contrat de maintenance nécessaire : peut se faire à distance par le service informatique.
- Verrouillage de l'accès aux images par mot de passe Peut se coupler avec un système de sécurité (alarme...)

Inconvénients de la vidéosurveillance IP

- Peu de choix de caméras (taille, forme...)
- Complexe : nombreux réglages, présence d'un informaticien indispensable.
- Images de moins bonne qualité qu'avec une installation analogique.
- Important s'il faut prévoir l'installation du réseau.

c- Passer de l'analogique au numérique : pas difficile

Si l'on est équipé d'un système de vidéosurveillance analogique et que l'on souhaite passer au numérique, cela ne nécessite pas forcément de remplacer tout le matériel.

Pour bénéficier des avantages du numérique, on peut utiliser une carte télé (à installer dans l'ordinateur) ou un magnétoscope numérique (DVR) qui peuvent convertir les signaux analogiques en signaux numériques : cela permet de stocker les images numérisées sur un ordinateur.

Grâce aux enregistreurs numériques, on peut facilement mettre en place un système de vidéosurveillance hybride qui associe caméras analogiques et caméras numériques ou IP via un réseau unique.

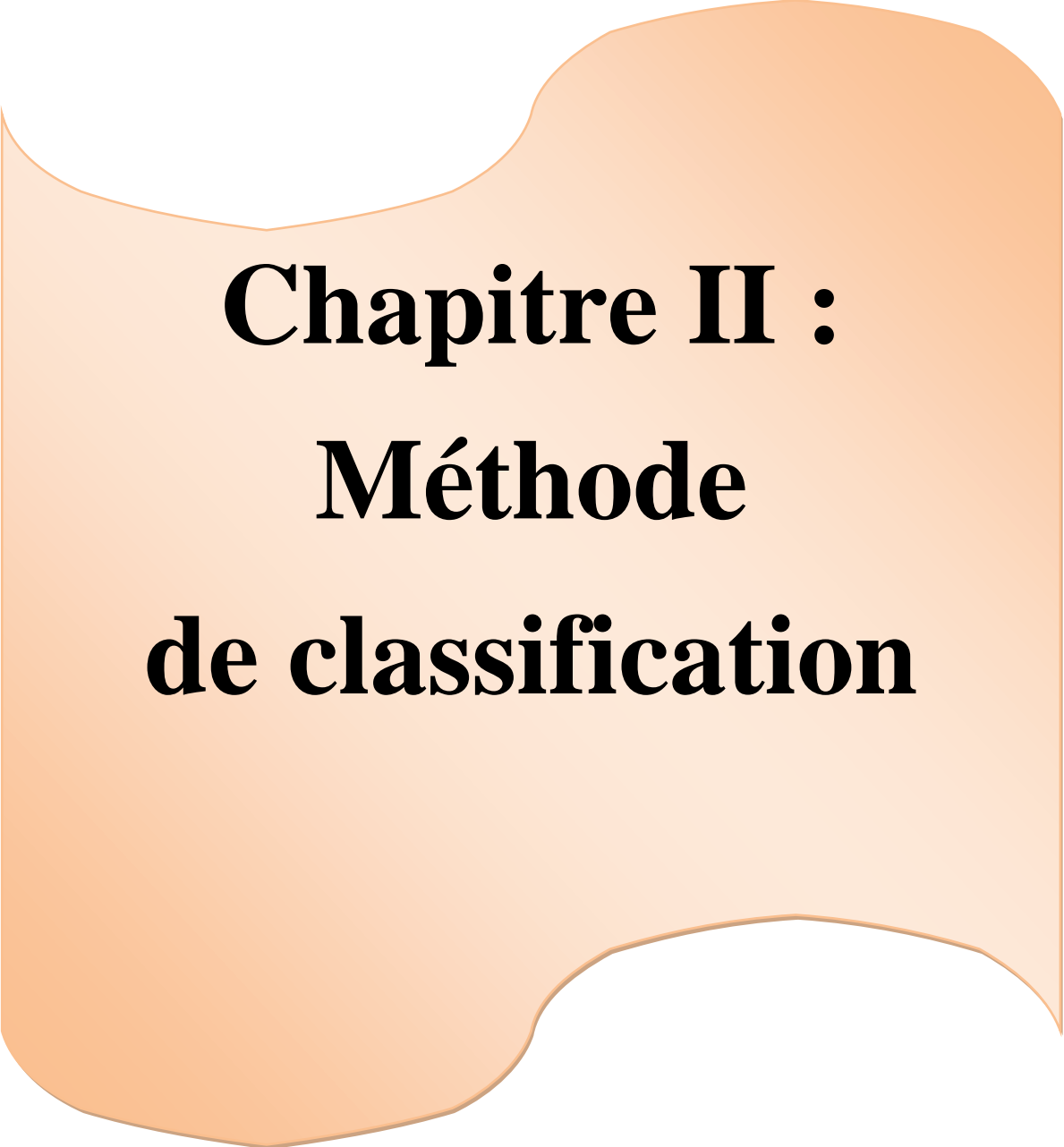
1.4.6. Domaines d'application de vidéosurveillance [10]

La vidéosurveillance, un champ d'application illimité :

- L'industrie
- Le transport privé ou collectif
- Le commerce et la distribution
- Les administrations et les services publics
- La santé
- Les lieux publics
- L'enseignement
- Les banques
- Les loisirs

1.5. Conclusion

On a présenté brièvement dans ce chapitre une définition d'image, de vidéo et de vidéo surveillance ainsi que ses types, ses avantages et ses inconvénients.

A large, light orange decorative shape with a wavy, irregular border, resembling a stylized book cover or a piece of paper. It is centered on a white background.

Chapitre II :
Méthode
de classification

2.1. Introduction

L'analyse des activités humaines, sur la base de séquences vidéo, nécessite différents niveaux de traitements. Les traitements de bas niveaux qui consistent en la détection des zones de mouvements pertinents. Les traitements de niveaux intermédiaires qui comprennent l'extraction d'information visuelle et leur représentation sous une forme concise et qui est la plus invariante possible. Enfin, les traitements de hauts niveaux permettant l'interprétation de ces informations et la reconnaissance de l'activité humaine. Il existe dans la littérature, une multitude de techniques pour mener à bien chacune de ces trois étapes.

Ces dernières années, les chercheurs ont de plus en plus recouru aux techniques d'apprentissage machine, notamment dans la détection de mouvements et soustraction de fond, Parmi ces techniques on a les Machines à Vecteurs de Support et les modèles de Markov cachés.

L'objectif de ce chapitre est de dresser un portrait général de ces outils et de son utilisation en apprentissage.

2.2. Les Machines à Vecteurs de Support (SVM)

2.2.1. Introduction

La méthode d'apprentissage génère une fonction qui fait correspondre à une image reçue en entrée une étiquette. Les caractéristiques jouent un rôle important dans la classification. C'est pour cela qu'il faut choisir celles qui peuvent discriminer au mieux une classe d'une autre. Différents types de caractéristiques peuvent être utilisés telles que : la couleur, la bordure, la texture, le flux optique, etc. Différentes méthodes d'apprentissage sont disponibles telles que les arbres de décision [11], les réseaux de neurones [12], AdaBoost (Adaptive Boosting) [13], ou les machines à vecteurs de support (SVM) [14].

Par la suite on expliquera la méthode SVM qu'on va l'utiliser dans notre travail.

2.2.2. Historique

Les séparateurs à vastes marges reposent sur deux idées clés :

- la notion de marge maximale
- la notion de fonction noyau.

Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVM.

✓ L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner, et en 1973 par Richard Duda et Peter Hart dans leur livre *Pattern Classification*. Les fondations théoriques des SVM ont été explorées par Vapnik et ses collègues dans les années 70 avec le développement de la Théorie de Vapnik-Chervonenkis..

✓ L'idée des fonctions noyaux n'est pas non plus nouvelle: le théorème de Mercer date de 1909, et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozoener.

Ce n'est toutefois qu'en 1992 que ces idées seront bien comprises et rassemblées par Boser, Guyon et Vapnik dans un article, qui est l'article fondateur des séparateurs à vaste marge.

L'idée des variables ressorts, qui permet de résoudre certaines limitations pratiques importantes, ne sera introduite qu'en 1995. À partir de cette date, qui correspond à la publication du livre de Vapnik, les SVM gagnent en popularité et sont utilisées dans de nombreuses applications[26].

2.2.3. Définition

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik, les SVM constituent la forme la plus connue. SVM est une méthode de classification binaire par apprentissage supervisé, elle fut introduite par Vapnik en 1995. Cette méthode est donc une alternative récente pour la classification. Cette méthode repose sur l'existence d'un classificateur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonction dites noyau (kernel) qui permettent une séparation optimale des données

Les SVM sont utilisées pour regrouper des données en plusieurs classes à l'aide des marges maximales de l'hyperplan qui séparent une classe de l'autre [15]. La marge de l'hyperplan, qui est maximisée, est définie par la distance entre l'hyperplan et les points de données les plus proches. Les points de données qui se trouvent à la limite de la marge de l'hyperplan sont appelés vecteurs. Pour la détections de personnes, deux classes sont utilisées : la classe personne (les échantillons positifs) et classe pas une personne (échantillons négatifs). A partir d'exemples d'entraînement annotés manuellement suivant ces deux classes, le calcul de l'hyperplan parmi une infinité d'hyperplans possibles est

effectué. SVM a été utilisé par Papageorgiou et al. [17] pour la détection de piétons et de visages dans les images.

Ils représentent une famille d'algorithmes d'apprentissage qui s'inspire de la théorie statistique de l'apprentissage de Vapnik.

Le SVM est un classifieur binaire, l'apprentissage par ce classifieur consiste à apprendre une fonction discriminante linéaire f à partir d'un jeu d'entrée constituée d'un certain nombre de couples entrées-sorties afin de pouvoir produire une sortie y étant donnée une nouvelle mesure. Ce type de classifieur tente de trouver une frontière de décision permettant de séparer linéairement les exemples de la première classe des exemples de la deuxième classe dans l'ensemble d'apprentissage [5].

2.2.4. Principe de fonctionnement

Le principe de ces méthodes est de trouver le meilleur hyper plan qui maximise la marge entre les exemples d'apprentissage et l'hyper plan.

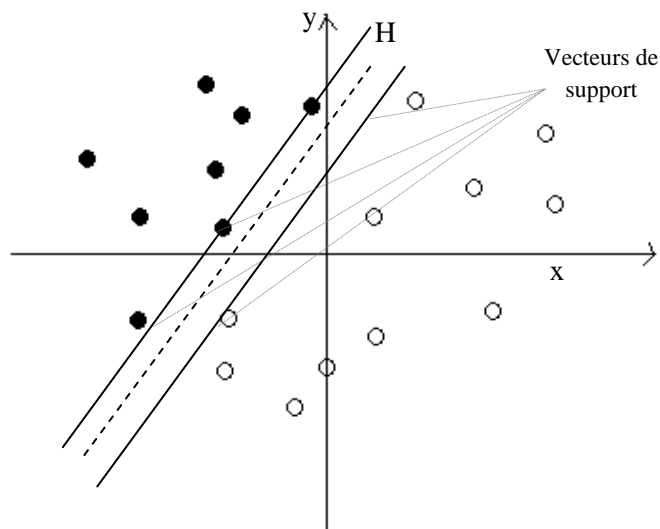


Figure 2.1 : Frontière de décision linéaire d'un classifieur SVM.

L'avantage principal de ces méthodes est qu'elles puissent être appliquées dans le cas où les classes ne sont pas linéairement séparables. Dans ce cas, les méthodes des SVM tentent de trouver des frontières de décision non linéaire. Le classifieur SVM emploie des fonctions de noyau (polynomial, gaussien) permettant de projeter les caractéristiques initiales dans un nouvel espace à grande dimension. Cette projection vise à rendre les données linéairement séparables dans le nouvel espace. Ensuite le classifieur SVM cherche

à trouver un séparateur linéaire dans le nouvel espace qui devient un séparateur non linéaire dans l'espace originale.

Les méthodes SVM ont été appliquées avec succès dans les problèmes de catégorisation des textes [18], de reconnaissances des actions humaines [19,20], de reconnaissance d'objets [21], etc.

2.2.5. Notions de base :

a- Hyperplan

Plaçons-nous dans le cas d'une classification binaire (i.e. les exemples à classer réparties en 2 classes). On appelle hyperplan séparateur un hyperplan qui sépare les deux classes, en particulier il sépare leurs points d'apprentissage. Comme il n'est en général pas possible d'en trouver un, on se contentera donc de chercher un hyperplan discriminant qui est une approximation au sens d'un critère à fixer (maximiser la distance entre ces deux classes) [27].

b- Vecteurs supports

Pour une tâche de détermination de l'hyperplan séparable des SVM est d'utiliser seulement les points les plus proches (i.e. les points de la frontière entre les deux classes des données) parmi l'ensemble total d'apprentissage, ces points sont appelés *vecteurs supports*.

c- Marge

Il existe une infinité d'hyperplans capable de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs supports), cette distance est appelée la marge.

2.2.6. Linéarité et non-linéarité [26]

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. Les premiers sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

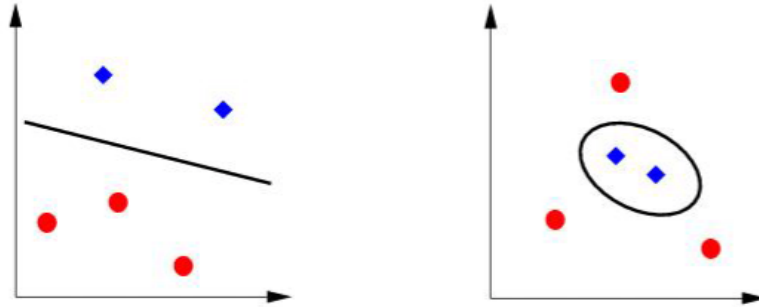


Figure 2.2 : Cas linéairement et non linéairement séparable.

2.2.7. Cas non linéaire :

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de re-description ». En effet, intuitivement, plus la dimension de l'espace de re-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant :

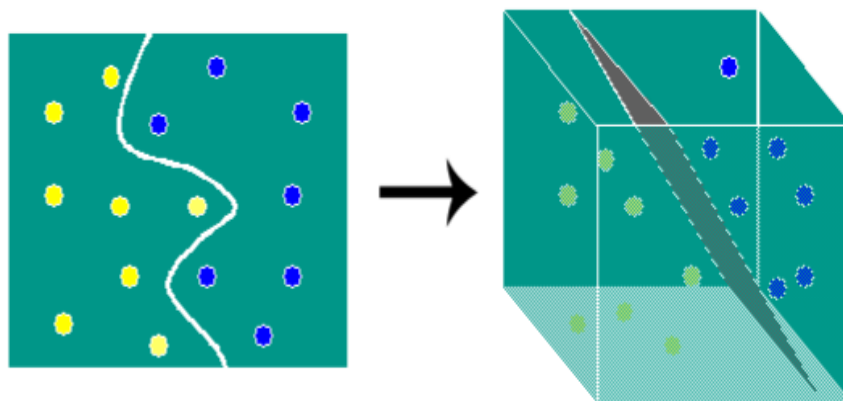


Figure 2.3 : Hyperplan dans le cas des données non linéairement séparable.

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. Cette transformation non linéaire est réalisée *via* une fonction noyau. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le

mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien.[26]

2.2.8. Noyaux SVM

Il existe plusieurs types de noyau :

- Noyau linéaire (simple produit scalaire) : $K(x ; x_i) = x * x_i$.
- Noyau Radial Basis Function (RBF) : $K(x ; x_i) = \exp(-\gamma \|x - x_i\|^2)$.
- Noyau Polynomial : $K(x ; x_i) = (x * x_i + c)^2$.
- Noyau sigmoïde : $K(x ; x_i) = \tanh(x * x_i + c)$.

Les noyaux RBF semblent donner les meilleurs résultats, dans le cas de la reconnaissance de chiffres manuscrits [27].

2.2.9. Avantage et Inconvénients

❖ Avantages

- Les SVM possèdent des fondements mathématiques solides.
- Les exemples de test sont comparés juste avec tous les exemples d'apprentissage.
- Décision rapide. La classification d'un nouvel exemple consiste à voir le signe de la fonction de décision $f(x)$ [16].

❖ Les limites

- Classification binaire d'où la nécessité d'utiliser l'approche un contre un.
- Grande quantité d'exemple en entrées implique un calcul matriciel important.
- Temps de calcul lors d'une régularisation des paramètres de la fonction noyau [29].

2.2.10. Domaines d'application d' SVM [28]

- ✓ Classification de données biologiques/physiques
- ✓ Classification de documents numériques
- ✓ Classification d'expressions faciales
- ✓ Classification de textures
- ✓ E-learning
- ✓ Détection d'intrusion
- ✓ Reconnaissance de la parole
- ✓ CBIR : Content Based Image Retrieval

2.3. Chaînes de Markov cachées

2.3.1. Introduction

La détection d'objet mobile en utilisant les modèles de Markov cachés ; MMCs, a pour but final de classer les pixels de chaque image incidente en deux régions : « objet mobile » et « Background ». Cela revient à déterminer la réalisation x des états cachés X , en partant de la seule information disponible représentée par la réalisation y de Y . Dans ce cas, on travaille dans un contexte dit « non supervisé ». Cela veut dire que les paramètres de ce modèle, sont estimés uniquement à partir de la réalisation y de l'observation Y , puisque la réalisation de X est totalement inconnue. Rappelons que ces paramètres s'arrangent en deux groupes : ceux propres à la chaîne de Markov cachée et ceux associés aux classes. Une fois ces paramètres estimés, on procède alors, à la détection.

Les modèles de Markov cachés (Hidden Markov Models ou HMMs) ont été introduits par Baum et al. à la fin des années 60. Ce modèle est fortement apparenté aux automates probabilistes, définis par une structure composée d'états et de transitions, et par un ensemble de distributions de probabilité sur les transitions. A chaque transition est associé un symbole d'un alphabet fini. Ce symbole est généré à chaque fois que la transition est empruntée [22].

2.3.2. Définition

Un MMC se définit également par une structure composée d'états et de transitions et par un ensemble des distributions de probabilité sur les transitions. La différence essentielle avec les automates probabilistes est que la génération de symboles s'effectue sur les états, et non sur les transitions. De plus, on associe à chaque état non pas un symbole, mais une distribution de probabilité sur les symboles de l'alphabet. Les modèles de Markov cachés sont utilisés pour modéliser des séquences d'observations. Ces observations peuvent être de nature discrète (par exemples les caractères d'un alphabet fini) ou continue (fréquence d'un signal, température) [22].

Un modèle de Markov caché (MMC) représente de la même façon qu'une chaîne de Markov un ensemble de séquences d'observations dont l'état de chaque observation n'est pas observé, mais associé à une fonction de densité de probabilité (fdp). Il s'agit donc d'un processus doublement stochastique, dans lequel les observations sont une fonction aléatoire de l'état et dont l'état change à chaque instant en fonction des probabilités de transition issues de l'état antérieur[23].

2.3.3. Processus stochastique

- Un processus stochastique (ou processus aléatoire) est une séquence X_1, X_2, \dots, X_n de variables aléatoires fondées sur le même ensemble fondamental S .
- Les valeurs possibles des variables aléatoires sont appelées les états possibles du processus.
- La variable X_t représente l'état du processus au temps t (on dit aussi l'observation au temps t). Les différentes variables aléatoires ne sont en général pas indépendantes les unes des autres. Ce qui fait réellement l'intérêt des processus stochastiques est la dépendance entre les variables aléatoires.
- Pour spécifier entièrement un processus stochastique, il suffit de spécifier:
 1. la loi de probabilité de la première variable aléatoire X_1 , qui spécifie donc l'état du processus lors de la première observation.
 2. pour toute valeur de $t > 1$ la probabilité conditionnelle:

$$P(X_{t+j} | X_1=i_1, \dots, X_{t-1}=i_{t-1})$$
 [25].

2.3.4. Présentation et applications des modèles de Markov cachés [22]

Nous nous limitons ici à la modélisation de séquences de symboles appartenant à un alphabet fini.

Un MMC est défini par un quintuple (S, Σ, T, G, π) , où :

- S est un ensemble de N états (représenter par w),
- Σ est un alphabet de M symboles (représenter par b_{ik}),
- $T = S \times S \rightarrow [0, 1]$ est la matrice de transition, indiquant les probabilités de transition d'un état à l'autre ; on note $P(s \rightarrow s_0)$ la probabilité de transition de l'état s vers l'état s_0 (représenter par a_{ij}),
- $G = S \times \Sigma \rightarrow [0, 1]$ est la matrice de génération, indiquant les probabilités de génération associées aux états ; on note $P(o | s)$ la probabilité de générer le symbole $o \in \Sigma$ à partir de l'état $s \in S$.
- $\pi : S \rightarrow [0, 1]$ est un vecteur de probabilités initiales de visite.

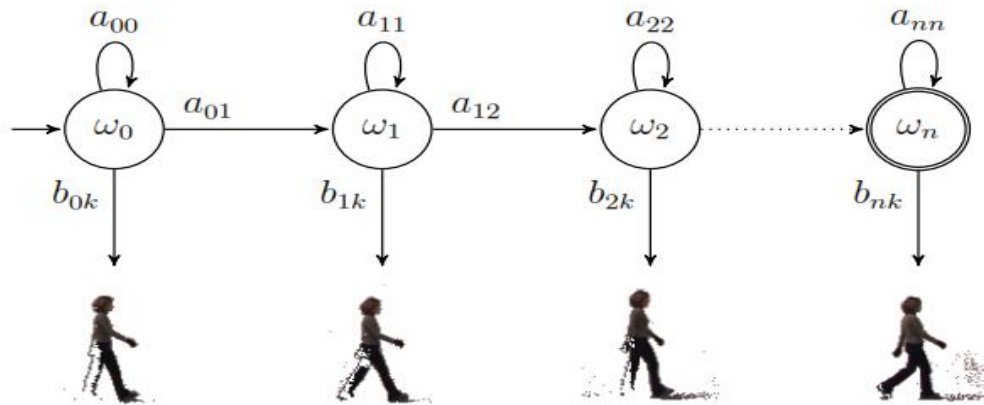


Figure 2.4 : Exemple de Modèle de Markov caché pour l'action «walk»[24]

La procédure de génération d'une séquence $o_1 \dots o_T$ de symboles à l'aide d'un MMC consiste à partir d'un état en suivant la distribution π , de se déplacer d'état en état suivant les probabilités de transition, et générer un symbole sur chaque état rencontré en utilisant la distribution de probabilité de génération associée à l'état. Lorsqu'un symbole a été généré, on choisit une transition sortante suivant la distribution de probabilité de transition associée à l'état courant, et la procédure est réitérée jusqu'à la $T^{\text{ème}}$ génération de symbole. Les MMCs définissent donc un processus stochastique non déterministe, ce qui explique le nom donné à ce modèle. On peut classer les principales applications des MMCs en deux catégories.

La première traite des problèmes de reconnaissance ou de classification, la seconde a trait aux problèmes de segmentation de séquences, c'est-à-dire au découpage d'une séquence en sous-séquences de différents types [22].

2.3.5. Les algorithmes de MMCs

a- Algorithme forward-backward

Soit $\alpha_t(s) = P(o_1 \dots o_t, s_t = s \mid H)$ la probabilité d'avoir généré la séquence $O = o_1 \dots o_t$ et d'être arrivé sur l'état s à l'instant t . Cette variable peut être calculée de manière inductive:

1. initialisation: $\alpha_1(s) = \pi(s) \cdot P(o_1 \mid s)$
2. induction: $\alpha_t(s) = (\sum_{s' \in S} \alpha_{t-1}(s') \cdot P(s' \rightarrow s)) P(o_t \mid s)$

Connaissant $\alpha_T(s)$ la probabilité d'avoir généré la séquence O et d'être arrivé sur s pour tout $s \in S$, le calcul de $P(O/H)$ est immédiat:

$$P(O|H) = \sum_{s \in S} \alpha_T(s) \quad (2.1)$$

La phase d'initialisation requiert une opération pour chaque état du HMM, donc au total $O(N)$ opérations. Pour la phase d'induction, pour chaque instant et chaque état, on réalise $O(N)$ opérations. Sommé sur l'ensemble des états et la totalité des instants, la phase d'induction requiert donc $O(N^2T)$ opérations.

Cet algorithme est appelé forward car l'induction est réalisée en avant : on calcule tout d'abord la probabilité de générer le premier symbole de la séquence, puis à chaque étape de l'induction on rajoute un symbole et on réitère la procédure jusqu'à ce que l'on ait calculé la probabilité de génération de la séquence entière. Un algorithme similaire, l'algorithme backward, peut être utilisé pour réaliser ce calcul à l'envers. On utilise alors la variable backward $\beta_t(s) = P(o_{t+1} \dots o_T / s_t = s, H)$ qui exprime la probabilité de générer la séquence $O = o_{t+1} \dots o_T$ en partant de l'état s . L'induction suit alors le schéma :

1. initialisation : $\beta_T(s) = 1$
2. induction : $\beta_t(s) = (\sum_{s' \in S} \beta_{t+1}(s') P(s \rightarrow s') P(o_{t+1} | s'))$

Connaissant la probabilité de générer la séquence O en partant de l'état s , le calcul de $P(O/H)$ peut alors être réalisée suivant la formule

$$P(O|H) = \left(\sum_{s \in S} \pi(s) \beta_1(s) \right) \quad (2.2)$$

avec la même complexité.

b- Algorithme de Viterbi

Soit $\delta_t(s) = \max_{s_1 \dots s_{t-1}} P(s_1 \dots s_t = s, o_1 \dots o_t | H)$ la probabilité maximale de générer la séquence $O = o_1 \dots o_t$ suivant un unique chemin arrivant sur l'état s à l'instant t . De la même manière que pour $\alpha_t(s)$, cette variable peut être calculée de manière inductive :

1. initialisation : $\delta_1(s) = \pi(s) P(o_1 | s)$
2. induction : $\delta_t(s) = \max_{s' \in S} (\delta_{t-1}(s') P(s' \rightarrow s)) P(o_t | s)$

Connaissant $\delta_T(s)$ pour tous les états s , on peut calculer la probabilité maximale $P(O/H, V)$ de générer O avec H suivant un simple chemin V par :

$$P(O | H, V) = \max_{s \in S} (\delta_T(s)) \quad (2.3)$$

En fait, ce n'est pas la valeur de cette probabilité qui nous intéresse mais réellement le chemin qui permet de générer O avec cette probabilité. On doit donc, à chaque étape t de l'induction et pour chaque état s, mémoriser l'état $s' = \psi_t(s)$ qui maximise l'équation d'induction. Alors :

1. initialisation : $\psi_1(s) = 0$
2. induction : $\psi_t(s) = \arg \max_{s' \in S} (\delta_{t-1}(s')P(s' \rightarrow s))$

Une fois les variables $\delta_t(s)$ et $\psi_t(s)$ calculées pour chaque étape de l'induction et pour chaque état, il ne reste plus qu'à lancer une procédure inductive de rétro-propagation pour "dérouter" le chemin de Viterbi $s_1^* \dots s_T^*$:

1. initialisation : $s_T^* = \arg \max_{s \in S} (\delta_T(s))$
2. induction : $s_t^* = \psi_{t+1}(s_{t+1}^*)$, $t \in \{T-1 \dots 1\}$

Mise à part la phase de rétro-propagation, l'algorithme de Viterbi est très similaire à l'algorithme forward. La principale différence résulte de la maximisation des probabilités attachées aux états précédents au lieu du calcul de la somme de ces probabilités.

c- Entraînement de Baum-Welch

L'algorithme d'entraînement de Baum-Welch est un algorithme qui cherche à estimer les paramètres $\lambda = \langle T, G, \pi \rangle$ du HMM en maximisant λ . C'est un algorithme d'estimation itératif qui fonctionne sur le même principe que l'algorithme d'entraînement de Viterbi [22]

2.3.6. Apprentissage d'un MMC

Les deux algorithmes précédents supposent disposer d'un MMC construit et paramétrée de manière à modéliser de façon satisfaisante les séquences à traiter. La question au centre de ce paragraphe est celle de la construction d'un tel MMC. Dans le cas le plus favorable, le MMC recherché peut être construit directement à partir des connaissances *a priori* dont on dispose sur le domaine. Dans la plupart des applications, le MMC doit être construit à l'aide d'un algorithme d'apprentissage.

Ces algorithmes sont appliqués sur un ensemble de séquences représentatives des séquences que l'on souhaite modéliser est appelées **séquences d'apprentissage**. On peut distinguer dans le problème de l'apprentissage d'un MMC deux cas de figure distincts,

suivant que la structure (nombre d'états du MMC et transitions autorisées) est connue ou ne l'est pas. Lorsque la structure est connue, le problème se réduit à un problème d'entraînement consistant à estimer les paramètres numériques (distributions de probabilité de première visite, de transition et de génération) de manière à expliquer au mieux les séquences d'apprentissage. Pour certaines applications, on ne dispose pas de connaissances suffisantes pour inférer naturellement la structure du MMC. L'apprentissage devient alors encore plus difficile. Il ne suffit plus de paramétrer une structure mais il faut également déduire cette structure des exemples fournis.

2.3.7. Différents types de MMCs

Quelques structures de modèles

Différentes topologies de MMC peuvent être envisagées, à commencer par le modèle ergodique, celui où chaque état peut être atteint à partir de n'importe quel autre en un nombre fini de transitions (**figure 2.5**).

Le modèle est donc caractérisé par une matrice T aux coefficients strictement positifs.

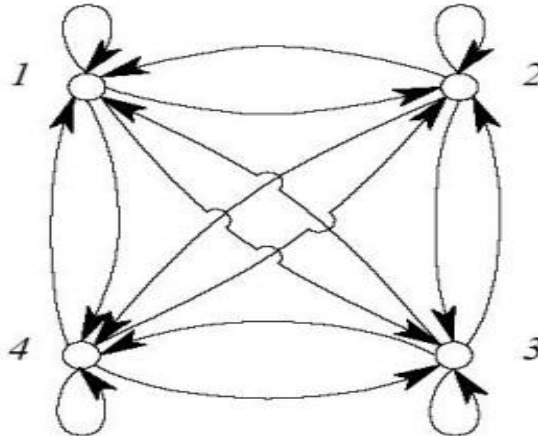


Figure 2.5: modèle de chaîne de Markov ergodique.

Une structure couramment utilisée est celle des modèles de **Bakis**, dits également modèles **gauche-droite** (**figure 2.6**), ainsi appelées parce qu'ils n'autorisent aucune transition d'un état vers un autre d'indice inférieur : les états qui se succèdent ont donc des indices égaux ou supérieurs aux précédents. Une fois dans le dernier état, le système est condamné à y rester : c'est pourquoi la probabilité initiale du premier état est posée égale à 1, les autres étant égales à 0.

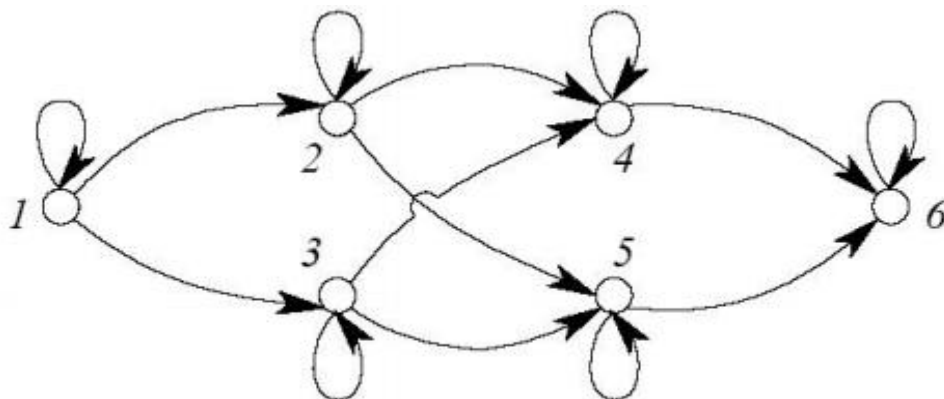


Figure 2.6 : modèle de Bakis.

Dans le même ordre d'idées, les sauts de nombreux états consécutifs sont souvent interdits ($T(i, j) = 0, j - i > \Delta$). Une variante du modèle gauche-droite est le modèle gauche-droite parallèle, utile, par exemple, en reconnaissance de la parole, où il permet de prendre en compte la possibilité de multiples prononciations d'un même mot.

Les procédures de ré-estimation des paramètres d'un modèle de Markov sont telles que, si certains de ces paramètres sont initialement mis à zéro, ils resteront nuls tout au long de la phase d'entraînement. Ceci permet donc de conserver les contraintes imposées à la structure du modèle. Dans le cas spécifique d'un modèle de Bakis, l'entraînement ne peut toutefois pas être réalisé à partir d'une seule séquence d'observations car, étant donné la nature du modèle, trop peu d'observations dans une séquence sont associées à un état particulier. Afin de disposer de données suffisantes pour l'estimation des paramètres, la procédure de ré-estimation doit donc être modifiée de façon à prendre en compte des séquences d'observations multiples. Bien sûr, il existe évidemment encore bien d'autres topologies de Modèles de Markov Cachés.

2.3.8. Les avantages et Les inconvénients :

a- Les avantages

- Base mathématique solide pour comprendre son fonctionnement.
- Variabilité de la forme.
- Aligement temporel incorporé systématiquement.
- Reconnaissance réalisée par un simple calcul de probabilité cumulé
- Décision global sans obligation d'utiliser des seuils
- Séparation french entre donnée et algorithme.

b- Les inconvénients


- Les choix a priori de la typologie des modèles (nombre d'état, transitions autorisées et règle de transitions).
- Dégradation des performances si l'apprentissage n'est pas suffisant. [22]

2.3.9. Domaine d'application

- ❖ Reconnaissance de formes et de la parole
- ❖ Traitement de la langue naturelle
- ❖ Commande automatique
- ❖ Traitement du signal
- ❖ Économie
- ❖ Robotique

2.4. Conclusion

Dans ce chapitre, nous avons présenté la méthode de classification SVM qui est utilisée dans le domaine de la vision par ordinateur. Nous avons aussi donné un aperçu sur les MMCs.

A large, light orange decorative shape with a wavy, irregular border, resembling a stylized book cover or a piece of paper. It is centered on a white background.

Chapitre III : Implémentation et expérience

3.1. Introduction

Dans de nombreuses applications, telles que la vidéo surveillance, la détection des comportements inhabituels ou la traduction en langue des signes, il est important de reconnaître l'activité de l'homme afin d'interpréter son comportement. Ce dernier peut être défini comme une succession temporelle d'actions primitives exécutées par le sujet humain le long du clip vidéo et qui peuvent être composées pour former une activité complexe.

Un système de reconnaissance d'action humain à partir des bases vidéo a pour objectif de satisfaire les besoins d'un utilisateur en sélectionnant les images les plus pertinents par rapport à ces besoins.

Nous allons donc en un premier temps mettre en œuvre une série de traitements qui seront appliqués à la vidéo en entrée afin d'optimiser les tâches subséquentes. Puis dans un second temps extraire des modèles représentant les mouvements à partir des séquences d'images afin de les soumettre au processus de reconnaissance et ainsi obtenir le mouvement décrit par celles-ci.

Dans ce chapitre, on va présenter l'implémentation de notre travail en présentant tous les étapes de conception de notre système nous allons aussi étudier expérimentalement les performances de la méthode de reconnaissance qu'on a proposée. Ensuite une explication des différentes tâches sera réalisées depuis le prétraitement, jusqu'à l'obtention d'une bonne reconnaissance.

Nous avons proposé, pour la reconnaissance des activités, une approche séquentielle fondée sur un modèle de classification hybride SVM-HMM prenant comme primitives d'entrée des descripteurs extraits des bases vidéos. Nous montrerons l'efficacité de notre modèle sur un environnement utile et une base de données qui seront décrites dans la première section. Dans la deuxième section, nous procéderons au choix des principaux paramètres de notre modèle de reconnaissance. Ensuite, dans la section trois, nous présenterons une évaluation de notre système global en le comparant avec les travaux de l'état de l'art. Enfin, une conclusion avec un bilan et perspective.

3.2. Environnement de programmation utilisée

Comme environnement de programmation, nous avons utilisé MATLAB R2013a pour implémenter notre système complet de reconnaissance d'action humaine, sous un système d'exploitation Windows7 Professionnel de type 64 bit. Et pour le matériel, nous utilisons un ordinateur portable de marque hp avec processeur Intel Core i3 de 2.10 GHz, possédant 4GO de RAM.

MATLAB permet le travail interactif soit en mode commande, soit en mode programmation, tout en ayant toujours la possibilité de faire des visualisations graphiques. Il possède les particularités suivantes :

- Puissance de calcul.
- la continuité parmi les valeurs entières, réelles et complexes.
- l'étendue de gamme des nombres et leurs précisions.
- la compréhension de la bibliothèque mathématique.
- l'inclusion des fonctions d'interface graphique et des utilitaires dans l'outil graphique.

Pour l'interface graphique, des représentations scientifiques et même artistiques des objets peuvent être créées sur l'écran en utilisant les expressions mathématiques ou bien directement en utilisant un outil graphique.

Les IHM (Interface Homme Machine), sont appelés GUI (Graphique User Interfaces) dans MATLAB. Elles permettent à des objets graphique (buttons, menus, cases à cocher,...) d'interagir avec un programme informatique [2].

Le jeu des expériences était extrait d'une base de données vidéo **KTH**.

3.3.La base de données

Dans cette section, nous présentons une description détaillée des bases de données utilisées pour valider notre système de reconnaissance. Nous présentons la base de données d'actions KTH, qui a été largement utilisée dans la littérature. L'ensemble de données, cependant, est composé d'actions simples avec un fond homogène. Relativement difficiles et vastes. Cette base de données offre relativement une grande variation de réalisation des actions humaines. Les actions sont réalisées d'une façon réaliste et sans aucune contrainte.

- Description de base de données KTH

La base de données KTH a été introduite par Schuldt . Cette base de données de faible résolution (images en niveau de gris de 160 _120 pixels) regroupe 6 types d'actions : marcher (walking), courir lentement (jogging), courir rapidement (running), boxer (boxing), faire un mouvement circulaire des bras (hand waving) et applaudir (clapping). Ces actions ont été effectuées plusieurs fois par 25 acteurs selon quatre scénarios : scènes en extérieur (s1), changement d'échelle (s2), changement de vêtements (s3) et scènes en intérieur (s4) (Figure 3.1). Dans la plupart des séquences, le fond est homogène ; toutefois, la présence de l'ombre peut rendre la soustraction de fond relativement difficile. En outre, il existe plusieurs facteurs de variations tels que les zooms pour le deuxième scénario, l'angle de prise de vue, les conditions d'éclairage et la durée des actions[30].

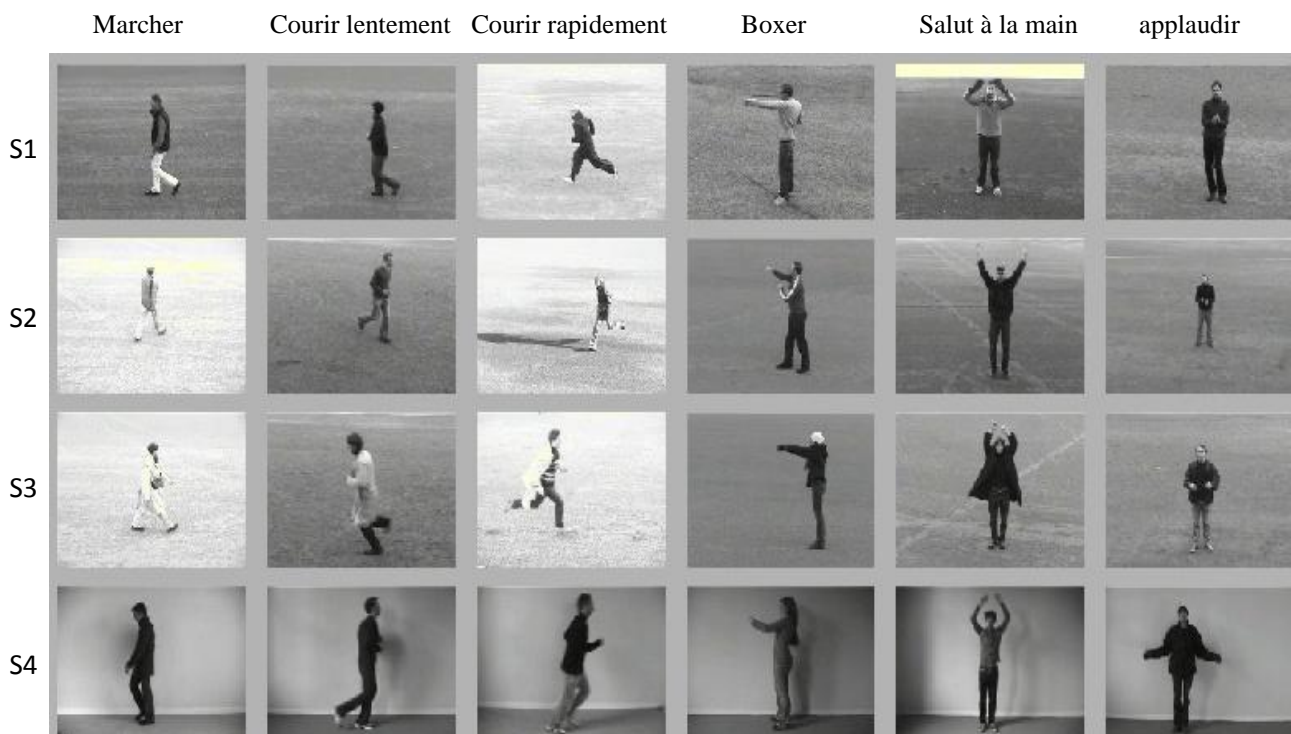


Figure 3.1 : Illustration de données KTH. Echantillon pour les six classes d'actions (colonne) enregistrées sous différents scénarios (par rangée).

3.4. Architecture du système proposé

Pour construire un système de reconnaissance d'action à partir de séquence vidéo, il est nécessaire de le composer en deux phases importantes : phase d'apprentissage et phase de reconnaissance.

Notre système de reconnaissance on av réalisé deux systèmes d'apprentissage, le premier se base sur SVM et le deuxième c'est un hybride entre SVM et HMM qui s'appuie sur le HMM comme classifieur de haut niveau ayant comme entrée les vecteurs de probabilités conditionnelles des classes fournis par le SVM. Ces classifieurs séquentiel permettent de prendre en considération l'aspect temporel des séquences vidéo tout en étant discriminant.

Notre modèle de reconnaissance offre également une grande flexibilité par rapport à l'ajout de nouvelles sources d'information encodant le contexte ou l'identité des objets manipulés.

a-Phase d'apprentissage SVM :

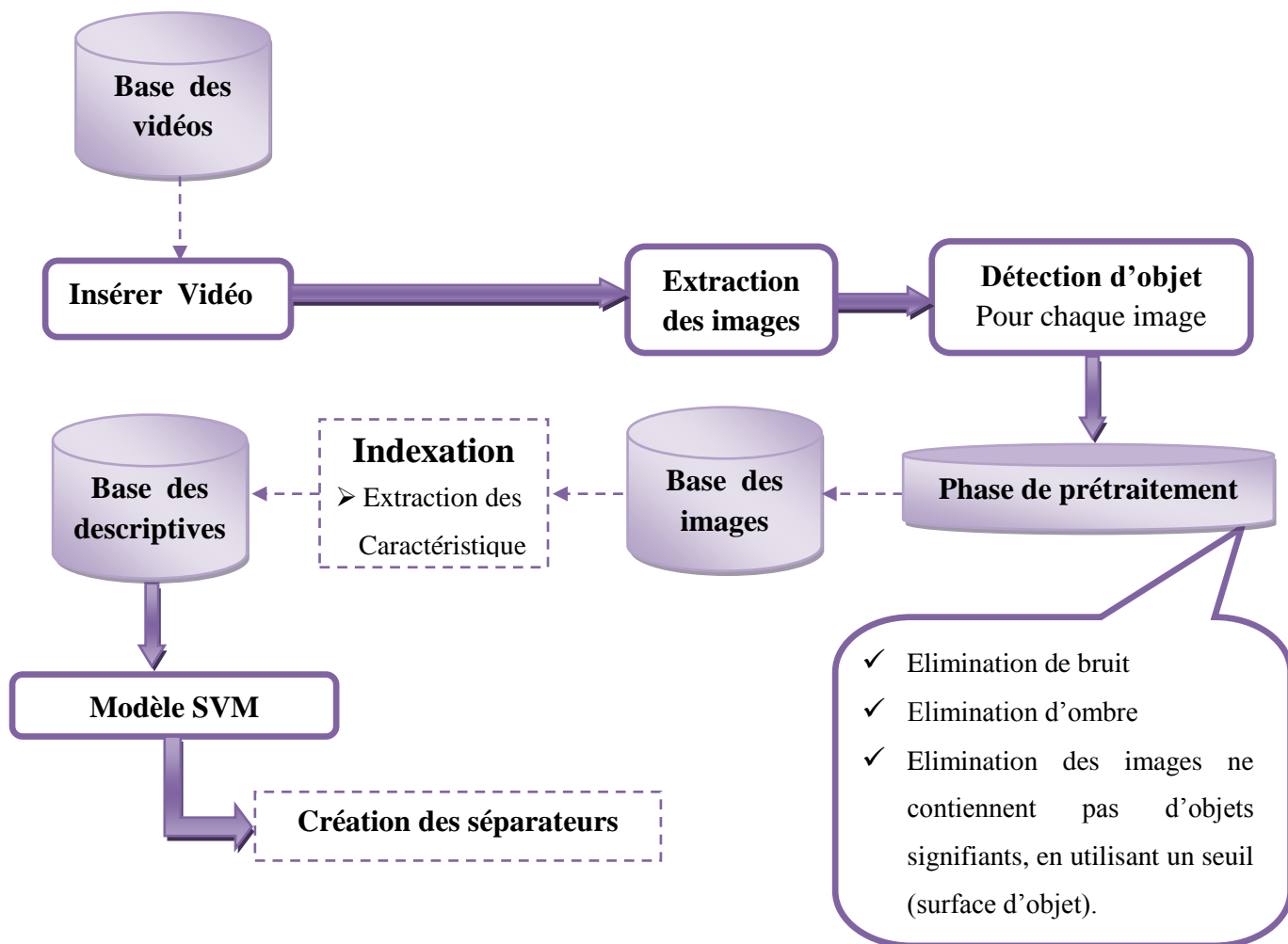


Figure 3.2: Structure de phase d'apprentissage SVM.

b- Phase d'apprentissage hybridation SVM-MMC :

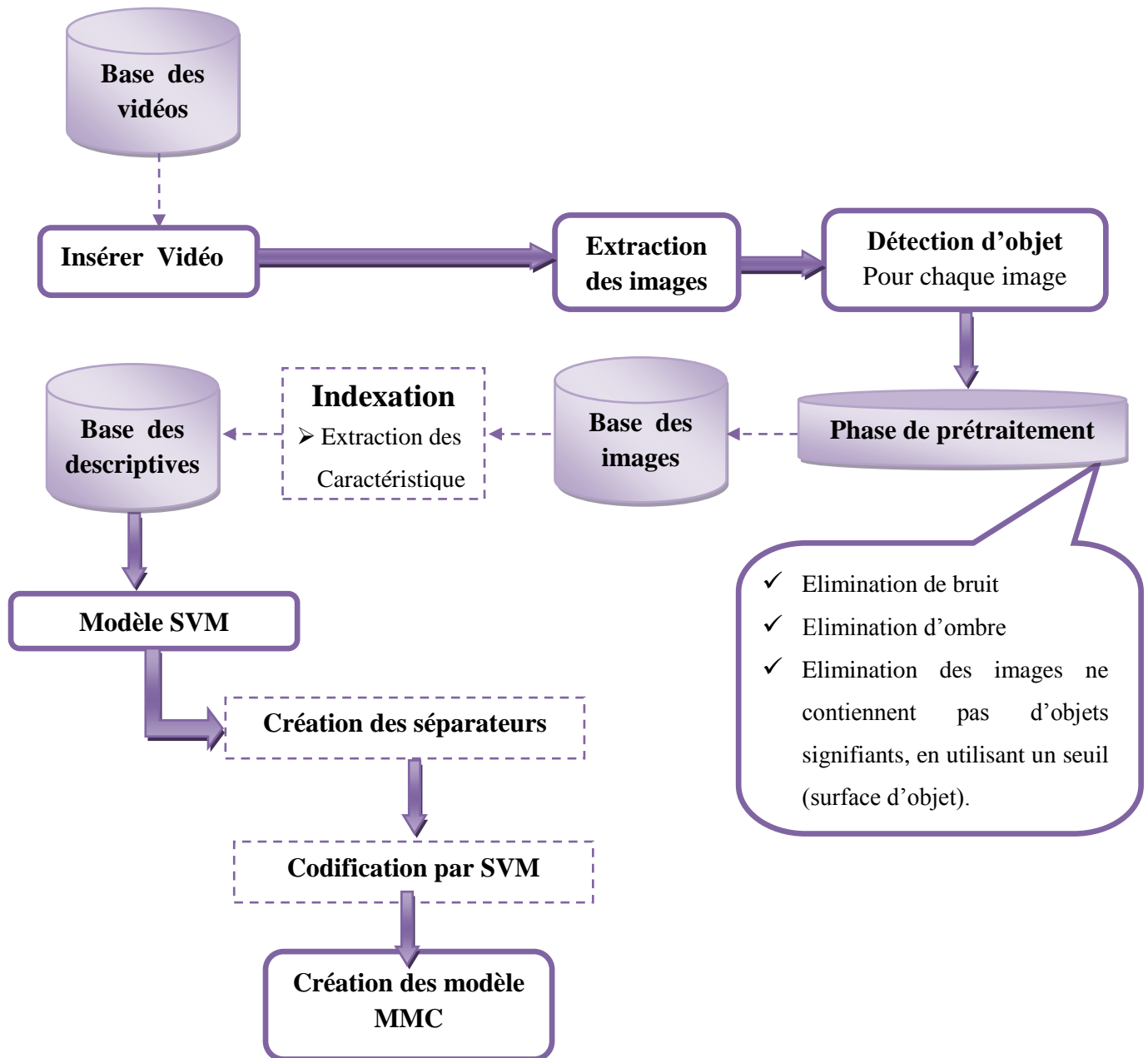


Figure 3.3: Structure de phase d'apprentissage pour le système hybride SVM-MMC.

c- Phase de reconnaissance :

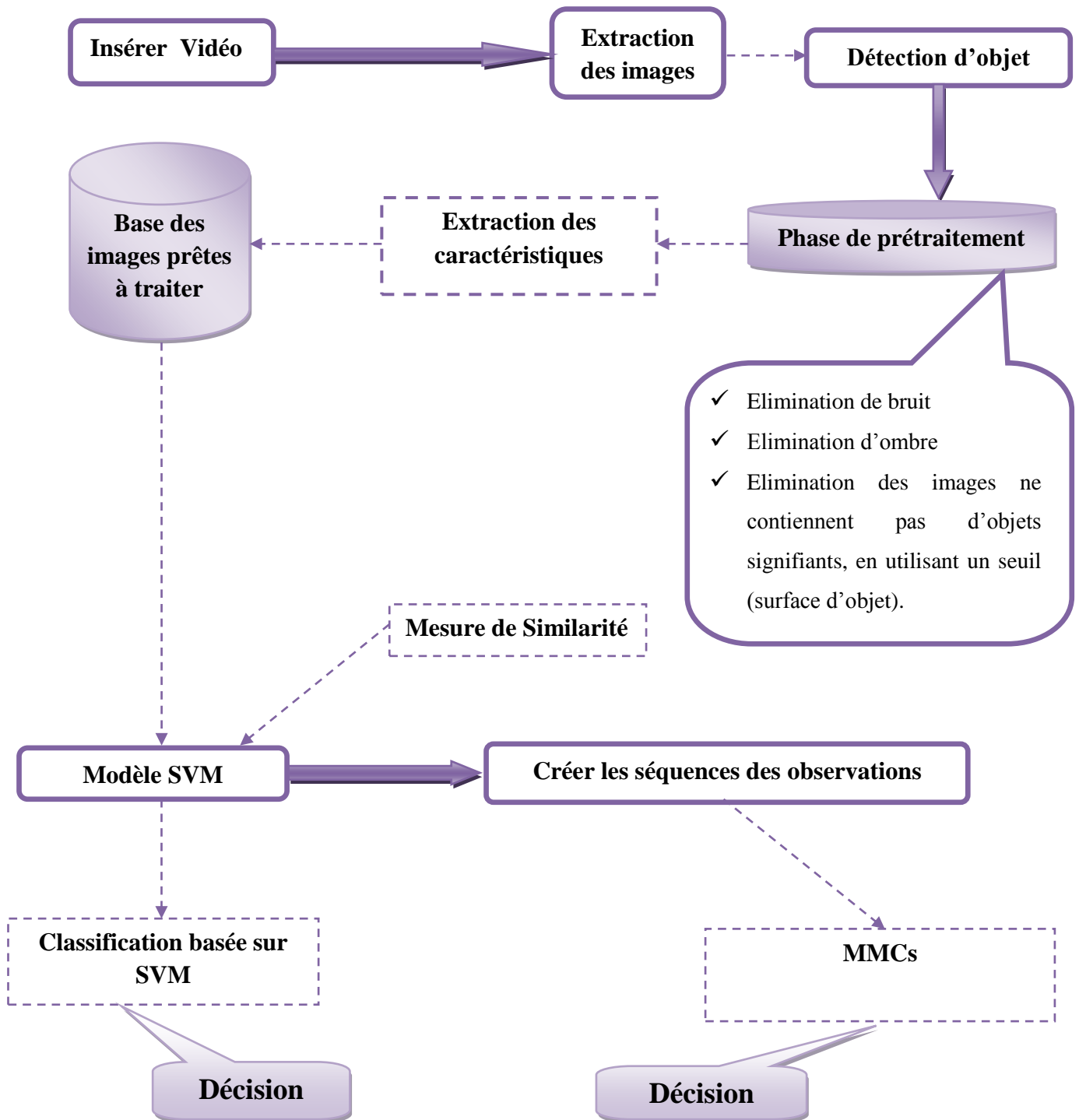


Figure 3.4: Structure de phase de reconnaissance pour SVM et système hybride SVM-MMC.

3.5. Explication détaillée des états du système

3.5.1. Lecture de séquences vidéo

Pour la première étape de notre système de reconnaissance, on va insérer les vidéos comme une unité d'entrer essentielle, pour extraire les images afin de faire la détection d'objet.

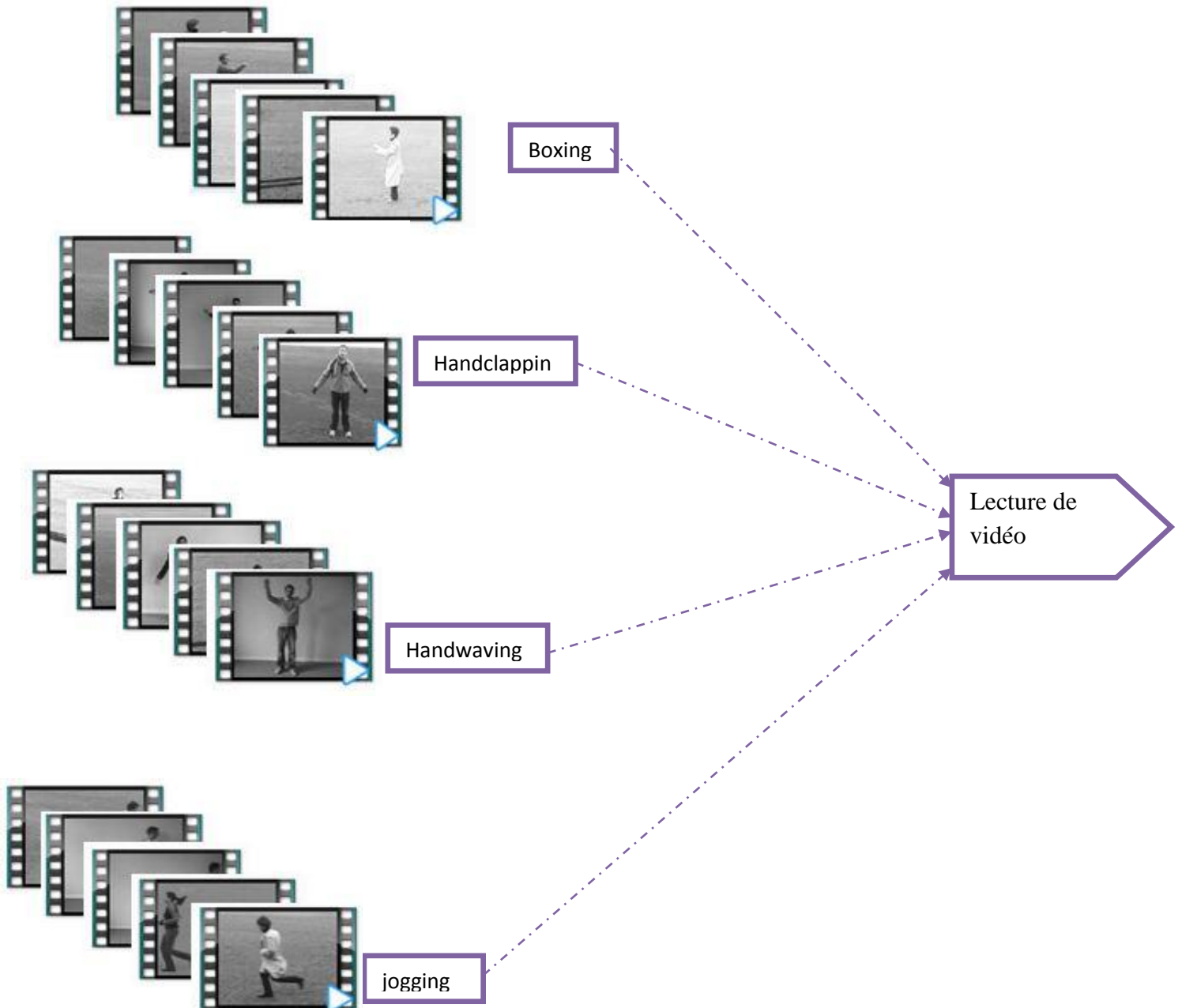


Figure 3.5: Représentation de lecture de chaque vidéo.

3.5.2. Phase de discrétisation de vidéo

On extrait un ensemble d'images du chaque vidéo, et on prend en considération le temps de discrétisation pour déterminer les objets mobiles.

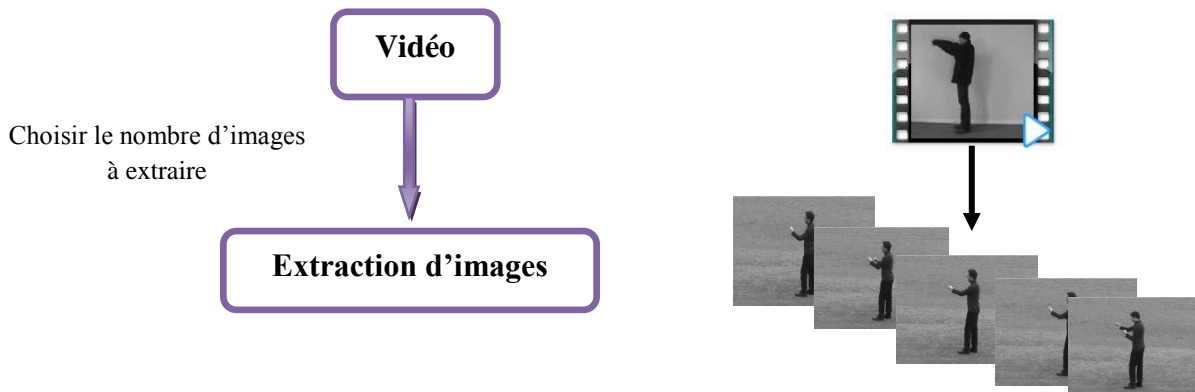


Figure 3.6: Représentation d'extraction des images.

➤ **Principe de nommage :** expliqué brièvement dans la figure ci-dessous ;

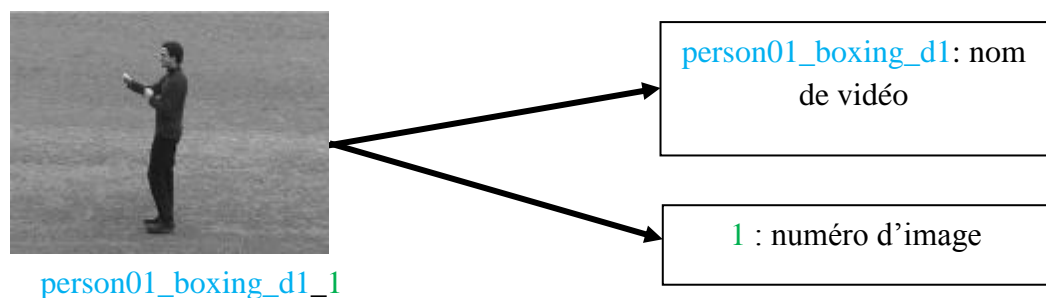


Figure 3.7: Manière de nommage des images.

3.5.3. Détection d'objet

On extraire l'objet du l'arrière plan, en utilisant l'algorithme suivante :

- 1- lire image
- 2- converti l'image RGB ver Gray
- 3- Appliqué la méthode de wolf pour l'extraction d'objet, (Wolf et al. [32] ont proposé de normaliser le contraste et le niveau de gris de l'image et de calculer le seuil en:

$$T = (1-K) \times m + K \times M + K \times \sigma / R (m-M)$$

Tels que K est fixé à 0,5, M est le niveau de gris minimum de l'image et R est l'écart type maximum de niveau de gris sur toutes les fenêtres obtenues [32].

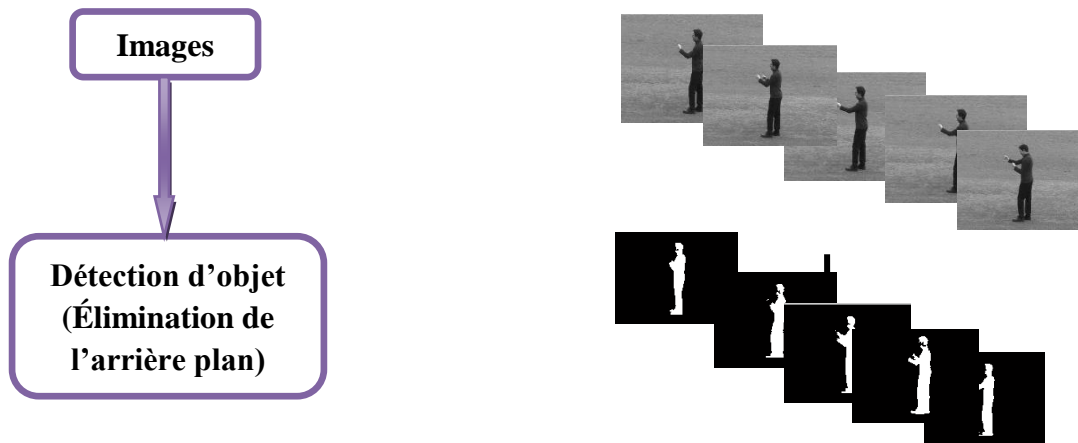


Figure 3.8: Méthode de détection d'objet

3.5.4. Phase de prétraitement

Pour l'élimination des objets indésirables on a utilisé un filtre moyenneur ; et pour éliminer l'ombre (**figure3.8**) on calcule la largeur de milieu pour déterminer un cercle aux pieds d'un rayon égale à ce largeur, on fait déplacer ce cercle verticalement jusqu'à le milieu de la forme en éliminant toutes les pixels trouvées hors les cercles.

Après l'extraction d'objet sans ombre, on fait une localisation des zones d'intérées qui déterminent les piets, les épaules et le milieu de la forme.

Par la suite une explication sera présentée de chaque étape avec des schémas pour facilitée la comprendre.

✓ Élimination d'ombre :

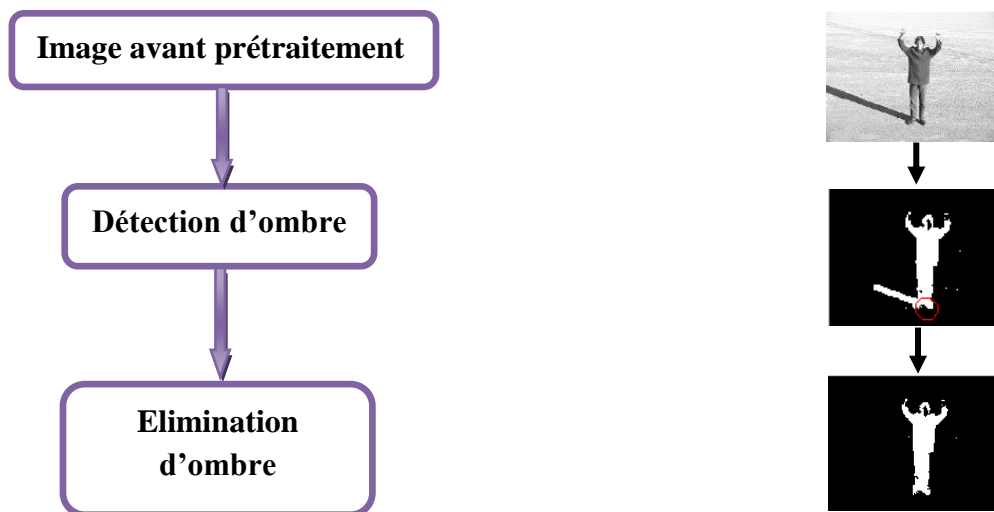


Figure 3.9: Méthode d'élimination d'ombre.

3.5.5. Mesure de similarité

Pour les caractéristiques utilisées, on a fait une étude pour quatre descripteurs géométrique **A**, **L1**, **L2** et **L3** (**figure3.10**), ils sont construits par la façon suivant :

Pour **L1** et **L2** on a divisé l'objet en deux zones, la première présente la partie haute c'est la zone tête de la personne, et la seconde présente la partie bas c'est la zone pied de la personne ; et on calcule **L1** c'est la ligne la plus large de la partie haut et **L2** c'est la ligne la plus large de la partie bas.

Et pour calculer **L3** c'est une régression linéaire des points de l'objet. Puis on conclut **A** c'est l'angle entre **L3** et l'axe de plan.

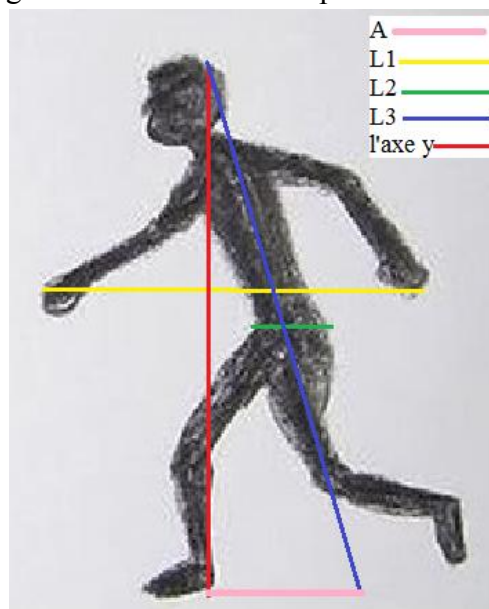


Figure 3.10: Explication des caractéristiques.

3.5.6. Modèle SVM

Pour notre choix de classification multiclasse SVM (**figure3.11**) on a utilisé l'approche "**un contre un**". Puis pour décider on a utilisé le graphe de décision DAGs.

3.5.7. Graphe de décision

Un graphe de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre). Pour notre approche nous utilisons DAGs (Directed Acyclic Graph).

C'est une méthode développée par Platt et all [33] pour résoudre le problème des zones d'indécision dans la méthode 1vs1. Premièrement, l'entraînement est effectué par la même méthode 1vs1 de la section précédente pour obtenir $K(K - 1)/2$ hyperplans. Puis, au lieu d'utiliser le vote pour l'affectation des nouveaux exemples, on construit un graphe de décision. Pour cela, on définit une mesure E_{ks} de la capacité de généralisation sur les différents hyperplans obtenus c.-à-d. pour chaque paire de classes. Cette mesure représente le rapport entre le nombre de vecteurs supports de l'hyperplan et le nombre d'exemples des deux classes correspondantes (équation 3.1).

Après la phase d'apprentissage on construit un graphe de décision qui sera utilisé pour la classification selon les étapes suivante :

- ✓ Créer une liste L contenant toutes les classes,
- ✓ Si L contient une seule classe, créer un nœud étiqueté de cette classe et arrêter.
- ✓ Calculer pour chaque paire de classes (i, j) la capacité de généralisation E_{ij} de l'hyperplan obtenu dans la phase d'entraînement 1vs1,
- ✓ Rechercher les deux classes k et s dont E_{ks} est maximum,
- ✓ Créer un nœud N du graphe étiqueté de (k, s).
- ✓ Créer un graphe de décision à partir de la liste $L - \{k\}$, de la même manière, et l'attacher au fils gauche de N,
- ✓ Créer un graphe de décision à partir de la liste $L - \{s\}$, de la même manière, et l'attacher au fils droit de N.

On obtient ainsi un graphe de décision similaire à l'exemple de la figure 3.10.

HC :handclapping	W: walking
HW : Handwaving	T: jogging
B : boxing	R: running

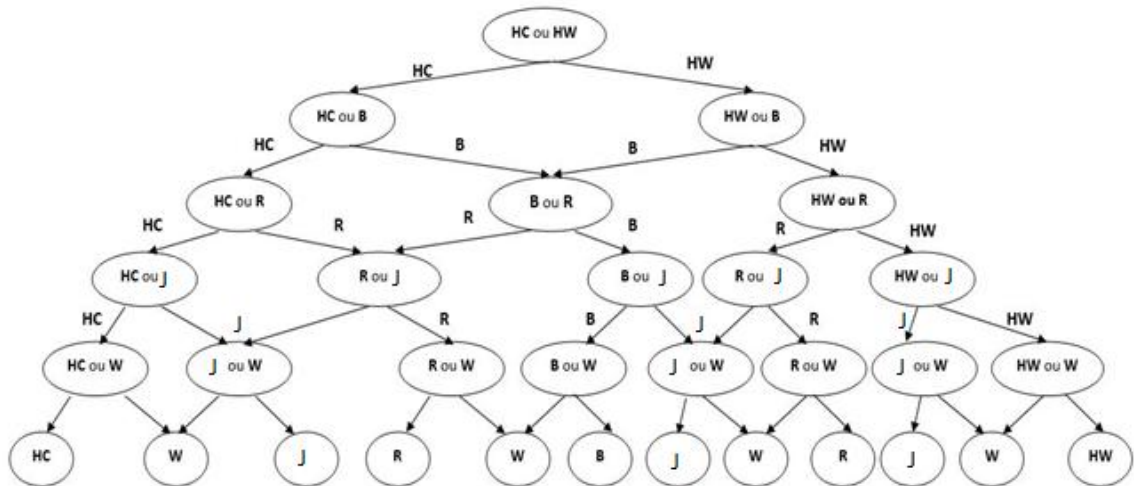


Figure 3.11: Structure de fonctionnement d'un SVM de type un contre un.

3.5.8. Construction des modèles MMCs

Après la construction de base de données des observations pour MMCs qui contient les résultats de SVM linéaire, on construit un modèle MMC pour chaque classe (action) (figure3.12).

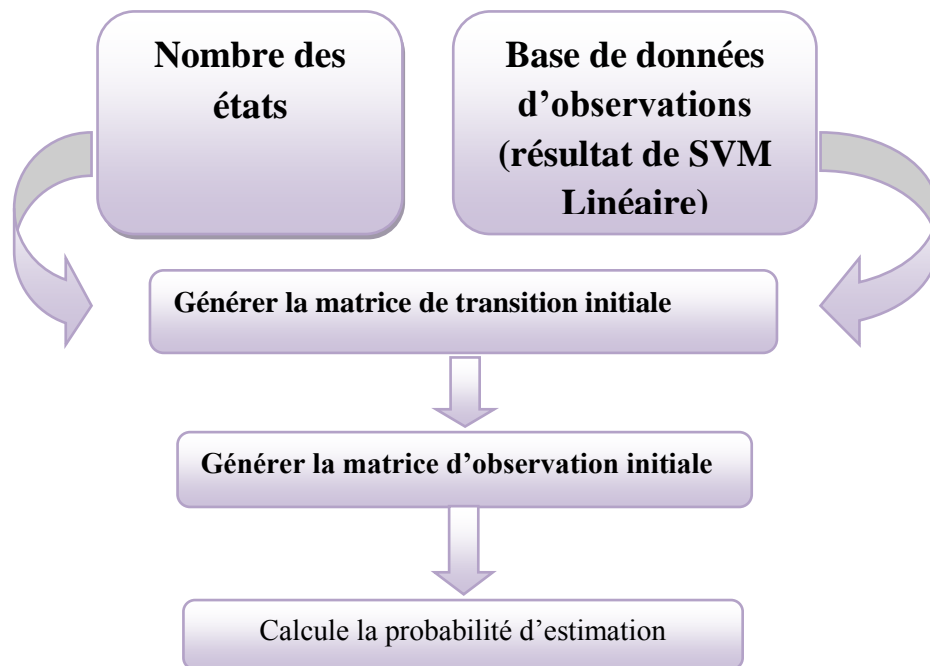


Figure 3.12: Structure de modèle MMCs.

Le schéma suivant explique le prétraitement

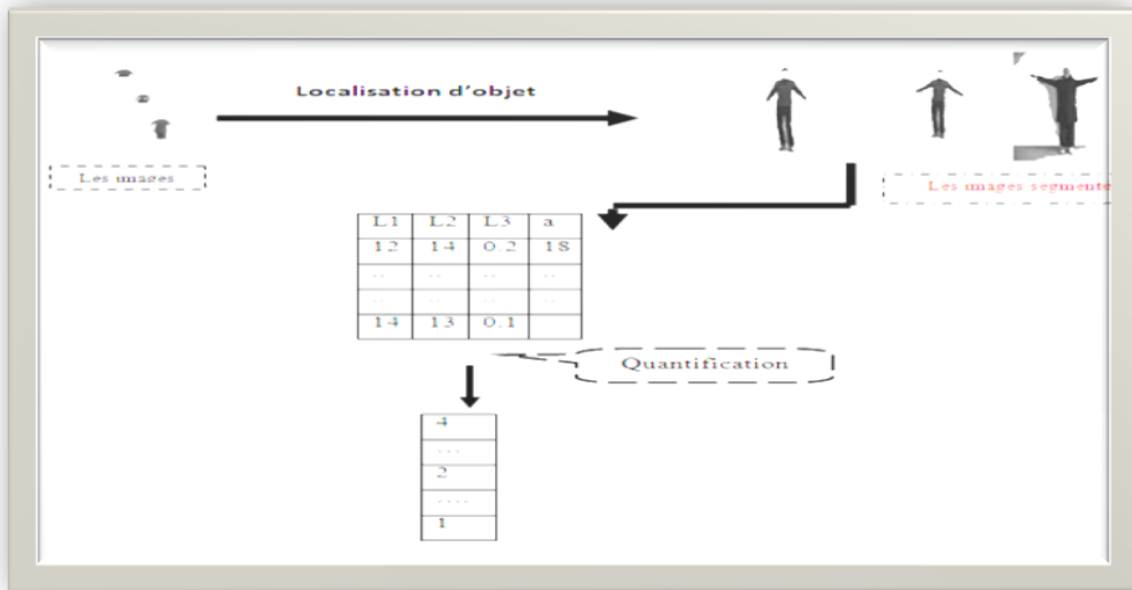


Figure 3.13: les étapes de phase de prétraitement.

3.5.9. Décision

Pour la phase de décision on fait le même traitement sur la vidéo donnée jusqu'à la phase d'apprentissage. Puis on choisie pour décider soit SVM ou le modèle hybride SVM-MMCs (figure3.14).

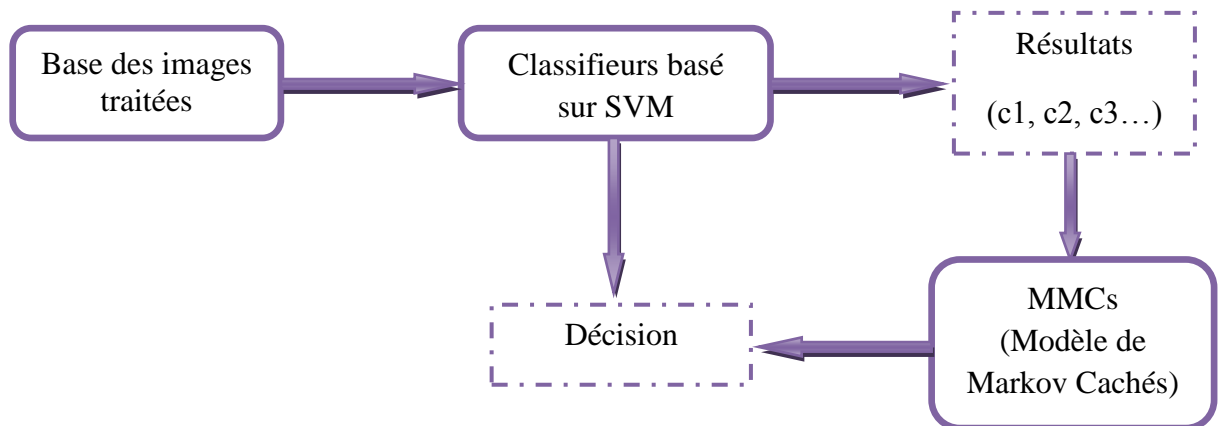


Figure 3.14: Schéma représente la méthode de prendre la décision.

3.6. Analyse des résultats

Dans ce qui suit, nous allons faire quelques testes pour avoir la performance de notre proposition et faire l'analyser. On a suivie deux stratégies pour faire ces testes :

3.6.1. La décision sur la vidéo globale

a- La méthode de classification avec SVM non linéaire

Pour SVM non linéaire on a utilisé Rbf (Radial bases function) comme kernel en variant sa seule paramètre gamma et on a des résultats semblables (**Tableau3.1** et **Tableau 3.2**), pour un taux de reconnaissance moyen de **43%**.

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	46.02	27.34	13.46	3.096	7.471	2.623	100
Handwaving	31.22	44.3	13.43	3.575	5.341	2.138	100
Boxing	11.62	11.29	51.29	7.306	14.55	3.952	100
Running	3.078	2.242	4.447	58.16	31.2	0.8724	100
Jogging	4.285	1.753	5.333	30.01	57.39	1.234	100
Walking	5.737	3.166	4.91	25.72	47.21	13.29	100
	102	90.08	92.87	127.9	163.2	24.11	

Tableau 3.1 : Matrice de confusion de SVM non linéaire kernel RBF avec $\gamma=0.5$

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	34.7	30.59	13.3	4.41	10.67	6.33	100
Handwaving	20.13	46.21	10.3	8.9	6.37	8.13	100
Boxing	13.8	4.71	41.1	14.91	15.31	10.17	100
Running	1.8	0.16	2.47	62.34	24.12	9.12	100
Jogging	2.45	2.81	2.99	38.46	51.2	2.1	100
Walking	4.31	2.4	4.81	36.6	34.7	17.2	100
	77.19	86.88	74.97	165.6	142.4	53.05	

Tableau 3.2: Matrice de confusion de SVM non linéaire kernel RBF avec $\gamma=1.5$

On présente dans le **tableau 3.1** et **tableau 3.2** la matrice de confusion obtenue en utilisant SVM avec un kernel RBF pour un $\gamma=0.5$ et un $\gamma=1.5$ pour la base KTH avec 6 classes. Les colonnes de cette matrice représentent les classes estimées, tandis que les lignes représentent les classes réelles.

On voit que les classes sont semblables sont :

La classe **handclapping** et la classe **handwaving**, aussi les deux classes **jogging** et **walking**.

La classe la plus séparable c'est la classe **running** avec un taux de reconnaissance de **62.34%**, suivie par la classe **boxing** avec un taux de reconnaissance de **51.29%**.

b- La méthode de classification avec SVM linéaire

Pour SVM linéaire on a un taux de reconnaissance = **35.19%**.

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	14.14	40.28	23.84	1.049	15.21	5.48	100
Handwaving	16.39	43.61	22.14	0.9318	11.53	5.407	100
Boxing	4.919	7.65	54.46	0.9374	25.26	6.773	100
Running	1.187	3.986	5.513	21.41	64.45	3.453	100
Jogging	0.6413	3.969	10.58	14.13	65.53	5.158	100
Walking	1.349	4.757	12.76	11.52	57.61	12.01	100
	38.62	104.3	129.3	49.97	239.6	38.28	

Tableau 3.3 : Matrice de confusion avec SVM linéaire

On présente dans le **tableau 3.3** la matrice de confusion obtenue en utilisant SVM linéaire, on voit que les classes sont semblables sont :

La classe **handclapping** et la classe **handwaving**, aussi les classes **walking** et **running** et **jogging**.

La classe la plus séparable c'est la classe **jogging** avec un taux de reconnaissance de **65.53%**, suivie par la classe **boxing** avec un taux de reconnaissance de **54.46%**.

c- La méthode de classification hybride SVM-MMCs

Pour notre proposition on a un taux de reconnaissance de **75.05%**.

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	67.68	19.19	1.01	0	12.12	0	100
Handwaving	39	41	1	0	19	0	100
Boxing	12.24	1.02	66.33	0	8.163	12.24	100
Running	0	0	0	91	9	0	100
Jogging	2	0	0	0	98	0	100
Walking	0	0	0	7	4	89	100
	120.9	61.21	68.34	98	150.3	101.2	

Tableau 3.4: Matrice de confusion avec hybridation SVM-MMCs

On présente dans le **tableau 3.4** la matrice de confusion obtenue en utilisant l'hybridation SVM & MMCs.

On voit qu'il y a une seule classe qui semble à l'autre; c'est la classe **handwaving** avec la classe **handclapping** avec un taux de reconnaissance assez acceptable de **41%**.

D'ailleurs les majorités des classes sont séparables, de classe **Jogging** avec un taux de reconnaissance de **98%**, suivie par la classe **running** avec un taux de reconnaissance de **91%**, ensuite la classe **walking** avec un taux de reconnaissance de **89%**, puis les deux classes **handclapping** et **boxing** avec un taux de **66%**.

3.6.2. La décision sur les frames

a. SVM non linéaire

Pour SVM non linéaire, comme on a mentionné précédemment on a utilisé Rbf (Radial bases function) ; les résultats sont représentées sous forme d'un matrice de confusion, on a obtenue un taux de reconnaissance de = **39.19%**.

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	40.73	20.7	13.74	5.126	9.943	9.759	100
Handwaving	25.06	38.89	14.31	5.369	8.27	8.105	100
Boxing	9.945	8.844	48.12	7.587	15.08	10.42	100
Running	3.453	1.744	10.92	41.96	24.17	17.75	100
Jogging	4.141	2.012	10.67	19.92	48	15.25	100
Walking	4.581	2.765	7.642	22.26	46.72	16.03	100
	87.91	74.96	105.4	102.2	152.2	77.33	

Tableau 3.5: Matrice de confusion de SVM non linéaire kernel RBF avec $\gamma=0.5$, nombre des images 10

On présente dans le **tableau 3.5** la matrice de confusion obtenue en utilisant SVM avec un kernel RBF pour un $\gamma=0.5$ avec un nombre d'image de 10 images ; pour la base KTH avec 6 classes. Les colonnes de cette matrice représentent les classes estimées, tandis que les lignes représentent les classes réelles.

On voit que les classes sont semblables sont :

La classe **handclapping** et la classe **handwaving**. Les classes les plus séparables sont la classe **boxing** la classe **jogging** avec un taux de reconnaissance de **48%**, suivie par la classe **running** avec un taux de reconnaissance de **41.96%**.

La classe **walking** a un mauvais taux de reconnaissance de **16.03%**

b. SVM linéaire

Pour SVM linéaire on a un taux de reconnaissance égal à **16.66%**, et un nombre des images de 10 images

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	8.527	6.95	5.393	49.51	10.47	19.15	100
Handwaving	8.527	6.95	5.393	49.51	10.47	19.15	100
Boxing	8.527	6.95	5.393	49.51	10.47	19.15	100
Running	8.527	6.95	5.393	49.51	10.47	19.15	100
Jogging	8.527	6.95	5.393	49.51	10.47	19.15	100
Walking	8.527	6.95	5.393	49.51	10.47	19.15	100
	51.16	41.7	32.36	297	62.84	114.9	

Tableau 3.6 : Matrice de confusion avec SVM linéaire avec frames, nombre d'images=10

On présente dans le **tableau 3.6** la matrice de confusion obtenue en utilisant SVM linéaire, on voit que les classes sont semblables sont :

La classe la plus séparable c'est la classe **running** avec un taux de reconnaissance de **49.51%**, suivie par la classe **walking** avec un taux de reconnaissance de **19.15%**, ensuite la classe **jogging** avec un taux de reconnaissance de **10.47%**.

c. SVM-MMCs

Pour notre proposition on a un taux de reconnaissance de **77.44%** avec un nombre d'image de 10 images.

	Handclapping	Handwaving	Boxing	Running	Jogging	Walking	
Handclapping	66.67	15.15	5.051	0	13.13	0	100
Handwaving	24	57	4	0	15	0	100
Boxing	11	4	63	0	12	10	100
Running	0	0	0	91	9	0	100
Jogging	2	0	0	0	98	0	100
Walking	0	0	0	8	3	89	100
	103.7	76.15	72.05	99	150.1	99	

Tableau 3.7: Matrice de confusion avec hybridation SVM-MMCs, en utilisant des frames de dix images

On présente dans le **tableau 3.7** la matrice de confusion obtenue en utilisant l'hybridation SVM & MMCs.

On voit que les majorités des classes sont séparables, de classe **Jogging** avec un taux de reconnaissance de **98%**, suivie par la classe **running** avec un taux de reconnaissance de **91%**, ensuite la classe **walking** avec un taux de reconnaissance de **89%**, puis les deux classes **handclapping** et **boxing** avec un taux de **63%**, finalement la classe **handwaving** avec un taux de reconnaissance de **57%**.

3.7. Etude comparative

Le tableau ci-dessous représente quelque approche de reconnaissance d'action humaine à partir de séquence vidéo ;

Méthode	Année	Taux de reconnaissance
Laptev et al. [Laptev et al., 2008]	2008	91.8 %
J. Yuan et al. [Yuan et al., 2009]	2009	93.3 %
Kovashka et al. [Kovashka and Grauman, 2010]	2010	94.5 %
Wang et al. [Wang et al., 2011a]	2011	94.2 %
Gilbert et al. [Gilbert et al., 2011]	2011	94.5 %
Zhang et al. [Zhang et al., 2012]	2012	94.1 %
Kaaniche et al. [Kaaniche and Bremond, 2012]	2012	94.7 %
Bilinski et al. [Bilinski et al., 2013]	2013	94.9 %
Notre modèle hybride	2017	77.44 %

Tableau 3.8: Comparaison des performances sur la base KTH.

Malgré ce résultat est un peu loin des autres mais en point de vue complexité notre méthode reste simple et on peut améliorer ces résultats en ajoutant d'autres caractéristiques.

3.8. Conclusion

Dans ce chapitre, on a présenté la méthode que nous avons adoptée pour reconnaître quelques actions humaines à partir d'une séquence vidéo en utilisant le classificateur SVM combiné avec les MMCs en basant sur les caractéristiques géométriques. La performance du système proposé a été évaluée avec la base de données KTH et a été comparée avec les résultats de certaines études précédentes.

Nous avons obtenu des résultats au moins aussi bons que ce que l'on trouve dans la littérature avec notre méthode.

A large, irregular orange shape with rounded corners and a wavy top and bottom edge, serving as a background for the text.

Conclusion générale

Conclusion générale

Conclusion générale

Dans ce mémoire, nous avons présenté notre contribution à la recherche concernant l'étude d'un système de reconnaissance automatique d'actions humaines par vision par ordinateur.

Nous avons décrit une nouvelle méthode pour la reconnaissance d'actions dans les vidéos basées sur une technique qui modélise des caractéristiques géométriques. Nous avons évalué cette méthode sur les bases vidéo KTH. On a développé deux méthodes de reconnaissance SVM ou hybridation SVM_MMCs. Nous obtenons des résultats au moins aussi bon que ce que l'on trouve dans la littérature.

A travers l'application que nous avons développée, nous avons montré que l'hybridation des méthodes de classification sert à améliorer les résultats.

Cependant, il est possible d'améliorer encore les résultats en développant de meilleurs détecteurs pour faire étudiée des caractéristiques plus spécifiques.

Finalement, L'avantage de notre système se manifeste par sa simplicité et son efficacité pour donner une meilleure représentation de l'action notre, ce travail présente des avantages certains en reconnaissance des actions humaines mais les possibilités d'extension sont encore très vastes, que ce soit au niveau des caractéristiques ou au niveau d'apprentissage.

A l'issue de ce travail, de nombreuses perspectives s'ouvrent sur les divers sujets traités. Les principales perspectives à cours terme incluent :

- Au niveau de phase d'apprentissage, on n'a peut pas traiter un grande nombre de vidéo.
- Au niveau de modélisation, notre méthode fondée sur le modèle SVM-HMM est extensible à la segmentation et la reconnaissance conjointes d'activités dans un flux vidéo continu. Une première investigation possible est d'appliquer les champs aléatoires conditionnels, en enteront le nombre des frames à traiter.
- Nous proposons aussi d'utiliser des bases plus grandes et difficiles à reconnaître.
- A plus long terme, le traitement des vidéos doit être online ; ce type d'application nécessite l'analyse d'une énorme quantité de données et requiert ainsi une grande capacité de calcul.

A large, light orange decorative shape with a wavy, irregular border, resembling a stylized book cover or a piece of paper. It is centered on the page and contains the text.

Références Bibliographiques

Références Bibliographiques

- [1] J-L. Baril "Outils Mathématique pour l'informatique" Université de Bourgogne Labo.le2i,UMR-CNRS 5158. 2007.
- [2] Guesdon VINCENT,"Détection efficace de contours d'images" ,université du Québec en Outaouais, Octobre 2004.
- [3] Y. Kabir " Segmentation d'images de films de radiographie dédiée au contrôle non destructif (CND) ", Thèse de magister, université de Saad Dahleb, Blida,1999.
- [4] Brogui M. & Boumaraf N. , (2015), "Detection et suivi d'un objet suspect dans le contexte de vidéosurveillance", Université Kasdi Merbah- Ouargla.
- [5] Touati R. , "Reconnaissance des actions humaines à partir d'une séquence vidéo", Mémoire pour l'obtention du grade de Maîtres sciences (M.Sc.) en informatique, Université de Montréal, Janvier 2014.
- [6] Le site <http://tpemultimedia.e-monsite.com/pages/qu-est-ce-que-la-video-surveillance.html>
- [7] Le site <http://www.inter-assistance.com/videosurveillance-guide.html>
- [8] Les différents matériels de vidéo surveillance.
- [9] Thi-Lan LE, université de NICE-SOPHIA ANTIPOLIS, " Indexation et recherche de vidéo pour la vidéo surveillance" , 3 Février 2009.
- [10] Le site <http://www.idvideo.com/solutions/videosurveillance/domaines-dapplications>
- [11] Lynne GREWE et Avinash C. KAK : Interactive learning of a multiple-attribute hash table classifier for fast object recognition. Computer Vision and Image Understanding (CVIU), 61(3):387–416, 1995.
- [12] Henry A. ROWLEY, Shume et BALUJA et Takeo KANADE : Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI), 20:23–38, 1996.
- [13] Paul VIOLA, Michael JONES et Daniel SNOW : Detecting pedestrians using patterns of motion and appearance. In International Conference on Computer Vision (ICCV), pages 734–741, 2003.
- [14] C.P. PAPAGEORGIOU, M. OREN et T. POGGIO : A general framework for object detection. In International Conference on Computer Vision (ICCV), pages 555– 562, 1998.

Références Bibliographique

- [15] Bernhard E. BOSER, Isabelle M. GUYON et Vladimir N. VAPNIK : A training algorithm for optimal margin classifiers. In 5th Annual ACM Workshop on Computational Learning Theory, pages 144–152, 1992.
- [16] Le site <http://www.ocean-securite.com/les-objectifs-et-le-fonctionnement-des-systemes-de-videosurveillance-c9-p22.html>
- [17] C.P. PAPAGEORGIOU, M. OREN et T. POGGIO : A general framework for object detection. In International Conference on Computer Vision (ICCV), pages 555–562, 1998.
- [18] J.T.Y. Kwok. Automated text categorization using support vector machine. Dans Proceedings of the International Conference on Neural Information Processing, ICONIP'98, pages 347–351, 1998.
- [19] S. Christian, L. Ivan, et C. Barbara. Recognizing human actions: A local svm approach. Dans Proceedings of the 17th International Conference on Pattern Recognition, ICPR'04, volume 3, pages 32–36, Washington, DC, USA, 2004. IEEE Computer Society.
- [20] H. Jhuang, T. Serre, L. Wolf, et T. Poggio. A biologically inspired system for action recognition. Dans IEEE Proceedings of the 11th International Conference on Computer Vision, ICCV'07, pages 1–8, Rio de Janeiro, Brazil, Octobre 2007. IEEE.
- [21] M. Pontil et A. Verri. Support vector machines for 3D object recognition. IEEE Trans. Pattern Anal. Mach. Intell., 20(6):637–646, Juin 1998.
- [22] Vincent Barra, "Apprentissage", Campus des Cézeaux - B.P. 1025 - 63173 AUBIERE CEDEX, 2005-2006.
- [23] Méthodes probabilistes – Modèles de Markov cachés.
- [24] BARNACHON M, "Reconnaissance d'action en temps réel à partir d'exemples", thèse de doctorat, 22 avril 2013.
- [25] Vincent L.& Ainane Y. , "Modèles de Markov cachés Reconnaissance de la parole", Faculté polytechnique de mons, 2004-2005.
- [26] M . Hasan et F. Boris, " SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges", BD Web, ISTY3, 16 janvier 2006.
- [27] C.L.Liu, K.Nakashima, H.Sako H.Fujisawa, Handwritten digit recognition using state-of-the-art techniques", IEEE, 2002.
- [28] CANU, S, "Machines à noyaux pour l'apprentissage statistique", 2007.

Références Bibliographique

- [29] BAKKI Mohamed, « Comparaison entre les technique de lecture labiale »,Mémoire de fin d'étude, en vue de l'obtention du diplôme de Master en informatique, Année Universitaire 201 3/2014 .
- [30] S.Mouna. 'Reconnaissance d'activités humaines _à partir de séquences vidéo'. Réseaux etTélécommunications [cs.NI]. Institut National des Télécommunications, 2014. Français. <NNT: 2014TELE0029>. <tel-01161610>
- [31] MESTAR Kh. , « Modèle de Markov Cachés », Université des Sciences et de la Technologie d'Oran, Département d'informatique, Option RFIA
- [32] C. Wolf, J. M. Jolion, and F. Chassaing. Extraction de texte dans des vidéos : le cas de la binarisation, Proceedings of 13^{ème} Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle, 2002, pp. 145 - 152.
- [33] J.C. Platt, N. Cristianini, and J. Shawe-Taylor. Advances in Neural Information Processing Systems, volume 12, chapter Large margin DAGs for multiclass classification, pages 547– 553. MIT Press, 2000.