

**République Algérienne Démocratique et Populaire**

**Ministère de l'enseignement supérieur et de la recherche scientifique**

**Université de Guelma**

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

**Mémoire de fin d'étude Master**



**Département d'Informatique**

**Spécialité : ingénierie des medias**

**Thème :**

---

---

***Recouvrement de posture 2D***

---

---

**Présenté par : Ayad Sawsen**

**Zeddouri Hanane**

**Sous la direction de :**

**Bourouaieh Douadi**

**Juin 2011**



# Remerciement

*En premier lieu et avant tous, on remercie énormément le GRAND DIEU Qui nous a aidé à réaliser ce travail.*

*Ensuite, nous tenons à exprimer notre profonde gratitude à notre encadreur Bourouaieh Douadi pour avoir accepté de nous suivre et qu'il nous a suggéré. Nous le remercions pour sa disponibilité, ses encouragements, son soutien moral, son expérience, sa sympathie, dont il nous a fait preuve tout au long de ce travail.*

*Nous remercions les membres du Jury d'avoir bien voulu accepter de faire partie de la commission d'examineur.*

*Nous voudrions adresser nos vifs remerciements à tous nos enseignants et personnels de département d'informatique qui ont contribué à notre formation.*

*Enfin, Nous remercions tous ceux qui ont participé, de près ou de loin, à la réalisions de ce mémoire, par leur savoir ou par leur soutien.*

**ZEDDOURI HANANE**

**&**

**AYAD SAWSEN**

---

---

## *Résumé*

---

---

*La reconnaissance de posture et mouvement à partir de séquence vidéo est un problème difficile. Ce problème s'inscrit dans le champ de recherche plus général de l'interprétation de séquence vidéo.*

*Durant ce mémoire, nous avons présenté un système de reconnaissance automatique des mouvements et poses en utilisant les modèles de Markov cachés probabilistes. Nous nous sommes intéressés à la reconnaissance des mouvements et postures d'un humain dans des scènes contraintes : mono caméra fixe, un seul sujet.*

*Le système comporte deux grandes phases. La première permet de concevoir la base d'apprentissage par un processus qui adopte plusieurs étapes. La deuxième cherche à reconnaître les mouvements et les postures.*

### *Les mots clés:*

Vision par ordinateur, Analyse du mouvement humain, Recouvrement de postures 2D, Chaines de Markov Caché(HMM), Quantification vectorielle, K-Means, Flot optique Open CV...etc.

---

---

## *Sommaire*

---

---

Introduction générale.....	1
----------------------------	---

---

---

### *Chapitre 1: La vision par ordinateur*

1. Introduction .....	4
2. Définition générale de la vision par ordinateur .....	4
3. Les approches en vision par ordinateur .....	5
4. Les Outils Fondamentaux de la Vision Par Ordinateur .....	5
4.1. Détection de Contours .....	6
4.2. Segmentation .....	8
4.2.1. Segmentation par classification.....	8
4.2.2. Segmentation par détection de contour.....	8
4.2.3. Segmentation par région.....	9
4.3. Calibrage de Caméra .....	10
4.4. Reconnaissance d'objet .....	12
4.5. Reconstruction 3D .....	12
5. Quelques applications concrètes .....	13
5.1. La surveillance intelligente .....	13
5.2. L'assistance et les interfaces homme-machine.....	14
5.3. Les jeux vidéo.....	14
6. Conclusion .....	14

---

---

### *Chapitre 2: Postures et mouvements humain*

1. Introduction .....	15
2. Grands domaines d'application.....	15
2.1. Surveillance Vidéo .....	15
2.2. Interface utilisateur avancée .....	16
2.3. L'identification et diagnostic basé sur le mouvement .....	16
3. Processus d'analyse du mouvement humain .....	17
4. La Détection .....	18
4.1. Segmentation de mouvement .....	18
4.1.1. Soustraction de fond .....	19

4.1.2. Méthodes statistiques .....	19
4.1.3. Différence inter-images .....	20
4.1.4. Flot optique.....	20
5. Le Suivi .....	21
5.1. Suivi à base de modèle .....	22
5.2. Suivi basé régions .....	23
5.3. Suivi basé caractéristiques .....	23
6. Estimation de pose .....	24
6.1. Sans modèle .....	24
6.1.1. Assemblage probabiliste de parties .....	24
6.1.2. Méthodes basées exemples .....	24
6.2. Utilisation indirecte de modèles .....	25
6.3. Utilisation directe de modèles .....	25
7. Reconnaissance .....	26
7.1. Interprétation de scènes .....	26
7.2. Approches holistiques de reconnaissance .....	26
7.3. Reconnaissance basée sur les parties du corps .....	27
8. Difficultés de la recherche et orientations futures .....	27
8.1. La Segmentation .....	27
8.2. La manipulation d'occlusion.....	28
8.3. L'utilisation de plusieurs caméras .....	28
9. Conclusion .....	28

---



---

### *Chapitre 3: Les Chaines de Markov Caché*

1. Introduction .....	29
2. Théorie des chaines de Markov .....	29
2.1. Chaîne Observable .....	30
2.2. Chaîne Cachée .....	31
2.2.1. Définition.....	31
2.2.2. Extensions des HMMs .....	33
2.2.2.1. Densités de probabilités.....	33
2.2.3. Différents types de HMM.....	35

2.2.3.1. Modèle ergodique .....	35
2.2.3.2. Modèle gauche-droite .....	35
2.2.4. Les trois problèmes fondamentaux pour les HMMs .....	36
2.2.4.1. Problème d'évaluation .....	36
2.2.4.2. Problème de reconnaissance .....	37
2.2.4.3. Problème d'apprentissage .....	37
2.2.5. L'utilisation des modèles de Markov cachés .....	38
3. La quantification vectorielle .....	38
3.1. Algorithme de quantification .....	39
4. Conclusion .....	41

---



---

### *Chapitre 4: Conception*

1. Introduction .....	42
2. Présentation générale .....	42
3. Partie d'apprentissage .....	43
3.1. Construction de la base des vidéos .....	44
3.2. Lecture de la vidéo .....	44
3.3. Extraction des vecteurs de caractéristiques .....	45
3.4. Quantification vectorielle (Classification des vecteurs de caractéristique par K-Means).....	47
3.5. Les résultats de la Classification (Quantification vectorielle) .....	48
3.6 L'apprentissage par les HMM (l'algorithme de Baum-Welch) .....	49
3.7. Les résultats d' HMM après l'entraînement (Baum-Welch) .....	50
4. Partie reconnaissance .....	50
4.1. La quantification vectorielle.....	52
4.2. La reconnaissance par les HMM (l'algorithme de Viterbi) .....	53
5. Conclusion .....	54

---



---

### *Chapitr5: L'implémentation*

1. Introduction .....	55
2. Les logiciels utilisés.....	55
2.1. Code Block .....	55

2.2. Open CV .....	56
3. Présentation de l'application .....	58
3.1. Base des vidéos .....	58
3.2. Extraction des vecteurs de caractéristiques.....	59
3.3. Quantification vectorielle .....	60
3.4. Les chaine de Markov caché.....	60
3.4.1. L'apprentissage.....	64
3.4.2. La reconnaissance .....	66
4. Les Résultats expérimentés .....	66
5. Conclusion .....	70
<hr/> <hr/>	
Conclusion générale.....	71

---

---

## Liste des figures

---

---

<b>Figure 1.1</b> : Exemple de détection de contours (Sobel).....	6
<b>Figure 1.2</b> : Calibrer une caméra consiste à estimer sa fonction de transfert.....	10
<b>Figure 1.3</b> : Exemple d'une séquence de 6 images d'une mire de calibrage. ....	12
<hr/> <hr/>	
<b>Figure 2.1</b> : Un cadre général d'analyse du mouvement humain .....	18
<hr/> <hr/>	
<b>Figure 3.1</b> : Graphe d'un Modèle de Markov Observable .....	30
<b>Figure 3.2</b> : Graphe d'un Modèle de Markov Caché.....	31
<b>Figure 3.3</b> : Modélisation HMM de l'exemple .....	33
<b>Figure 3.4</b> : Représentation d'un HMM discret à 3 états .....	34
<b>Figure 3.5</b> : Modèle de chaîne de Markov ergodique .....	35
<b>Figure 3.6</b> : Modèle de Bakis .....	35
<b>Figure 3.7</b> : Quantification vectorielle, $k=2, L=13$ .....	39
<b>Figure 3.8</b> : K-means .....	41
<hr/> <hr/>	
<b>Figure 4.1</b> : Vue générale de l'implémentation .....	43
<b>Figure 4.2</b> : grandes étapes de la phase apprentissage des HMM.....	44
<b>Figure 4.3</b> : Transport de la luminosité .....	45
<b>Figure 4.4</b> : le flot optique d'une personne qui marche.....	46
<b>Figure 4.5</b> : Un exemple de notre processus de calcul de flot optique .....	46
<b>Figure 4.6</b> : l'histogramme d'une frame du mouvement marche.....	47
<b>Figure 4.7</b> : Exemple sur la quantification par k_means .....	48
<b>Figure 4.8</b> : Le processus de la Classification .....	49
<b>Figure 4.9</b> : l'algorithme de Baum-Welch .....	50
<b>Figure 4.10</b> : les probabilités de transition d'un HMM (a) avant et (b) après entraînement....	50
<b>Figure 4.11</b> : Le processus de la reconnaissance de mouvement .....	51

<b>Figure 4.12</b> : le principe de la distance euclidienne .....	52
<b>Figure 4.13</b> : Les entrées et les sorties de l'algorithme de viterbi .....	53
<hr/> <hr/>	
<b>Figure 5.1</b> : l'interface de code block .....	55
<b>Figure 5.2</b> : Vue d'ensemble d'opencv .....	58
<b>Figure 5.3</b> : frame d'une vidéo de la marche.....	58
<b>Figure 5.4</b> : frame d'une vidéo de saut .....	58
<b>Figure 5.5 (a)</b> Frame18 d'une personne marche .....	59
<b>Figure 5.5 (b)</b> Frame12 d'une personne saute.....	59
<b>Figure 5.6</b> : L'histogramme des frames entre 0 et 8 .....	61
<b>Figure 5.7</b> : L'histogramme des frames entre 9 et 14 .....	62
<b>Figure 5.8</b> : L'histogramme des frames entre 15 et 23 .....	63
<b>Figure 5.9</b> : L'histogramme des frames entre 27 et 31 .....	63
<b>Figure 5.10</b> : Les probabilités de transition d'un HMM (poses)(a) avant et (b) après l'entraînement .....	64
<b>Figure 5.11</b> : Les probabilités de transition d'un HMM (mouvement globale) (a) avant et (b) après l'entraînement .....	65
<b>Figure 5.12</b> : Les probabilités de transition d'un HMM (a) avant et (b) après l'entraînement	66
<b>Figure 5.13</b> : Les résultats de l'HMM1 de pose sur les vidéos de la marche .....	67
<b>Figure 5.14</b> : Les résultats de l'HMM1 de mouvement sur les vidéos de la marche .....	67
<b>Figure 5.15</b> : Les résultats de l'HMM1 de pose sur les vidéos de saut .....	68
<b>Figure 5.16</b> : Les résultats de l'HMM1 de mouvement sur les vidéos de saut .....	68
<b>Figure 5.17</b> : Les résultats de l'HMM2 de mouvement sur les vidéos de la marche .....	69
<b>Figure 5.18</b> : Les résultats de l'HMM2 de mouvement sur les vidéos de saut .....	69
<hr/> <hr/>	

---

---

## *Liste des tableaux*

---

---

<b>Tableau 5.1</b> : Les vecteurs de caractéristique des quelque Frames de la vidéo.....	59
<b>Tableau 5.2</b> : les centres des classes .....	60

---

---

## *Introduction générale*

---

---

Ces dernières années, le problème de la reconnaissance des postures et des activités humaines a suscité l'intérêt de communautés de recherche de plus en plus larges.

Dans différentes applications, telles la vidéo surveillance, l'archivage ou l'indexation de vidéos, il est important de reconnaître les mouvements des personnes pour pouvoir interpréter leurs comportements.

La reconnaissance d'activité nécessite l'extraction de données multiples, l'interprétation automatique des séquences vidéo, et fait appel à des techniques d'analyse vidéo (perception visuelle, estimation de mouvement...) et des méthodes d'analyse et de classification de données. Ce problème d'identification devient crucial lorsque:

- Le nombre d'individus augmente,
- les points de vue de caméras sont différents,
- Ou encore lorsque les sujets sont dans des environnements complexes,

Concernant les postures, leur nombre dépend du degré de liberté du corps humain (les articulations comme les épaules ou les genoux). En outre, la morphologie de la personne (taille, corpulence, etc...) influence la perception de la posture. Enfin, les vêtements peuvent aussi donner différents types d'apparences pour la même posture.

Plusieurs méthodes utilisent l'historique du mouvement comme base pour la reconnaissance de gestes. Des caractéristiques appropriées sont extraites de l'image et différentes méthodes de classification sont utilisées tels que les réseaux de neurones et les modèles de Markov cachés probabilistes.

Certaines méthodes sont basées sur les automates et fournissent généralement une solution au problème de changement de phase avec l'utilisation de modèles probabilistes tels que les HMM. Une approche commune est de définir chaque caractéristique statique d'une action comme un état et d'apprendre la relation entre l'ensemble des caractéristiques. Pour classer

une action, la probabilité conjointe à la valeur maximale est retenue comme critère de classification.

*Le travail présenté dans ce mémoire s'insère dans le grand domaine de la vision par ordinateur qu'est l'analyse du mouvement. Dans notre cas nous considérons que les postures humaines sont déductibles à partir de la reconnaissance globale du mouvement. Par exemple, une personne qui effectue un saut passe par un ensemble de postures typiques.*

*Durant ce mémoire nous avons proposé une approche générique et fonctionnelle pour suivre les mouvements d'une personne, reconnaître ses postures et ses activités filmée par une seule caméra fixe.*

*Notre méthode se divise en deux parties : la première partie concerne l'apprentissage, cela est fait, premièrement, par la création d'une base des vidéos, puis la lecture de ces vidéos comme données, et l'extraction des vecteurs de caractéristiques par le calcul de l'histogramme de flot optique. L'étude des histogrammes permet de construire les HMM adéquats et de les initialiser. Ensuite, une quantification vectorielle se fait en utilisant la méthode du K-MEANS. Et enfin l'apprentissage est réalisé par l'un des algorithmes des modèles de Markov cachés qui est l'algorithme de Baum-Welch.*

*La deuxième partie concerne la reconnaissance, elle est conçue pour reconnaître les mouvements d'un humain en intégrant le résultat de la première partie. Cela est fait par le passage sur les mêmes premières étapes de l'apprentissage qui sont : extraction des vecteurs de caractéristiques et quantification vectorielle. La quantification vectorielle est faite par le calcul de la distance euclidienne des centres des classes données par l'application du K-MEANS. Enfin la reconnaissance est réalisée par l'algorithme de VITERBI qui permet de reconnaître les séquences d'états optimales (plus probable) dans un contexte d'états cachés et d'observations.*

*Ce mémoire est organisé comme suit:*

- *Chapitre 1* : présente une introduction générale sur le domaine de la vision par ordinateur, en incluant quelques outils fondamentaux de ce domaine et quelques applications concrètes.
- *Chapitre 2* : aborde un aperçu général sur la reconnaissance de postures et de mouvement humain, incluant le processus d'analyse de mouvement (la détection, le suivi, l'estimation de pose et la reconnaissance de posture)

- **Chapitre 3** : présente les modèles de Markov cachés HMMs. incluant une définition, les types d'HMMs, les éléments qui constituent un HMMs, les trois problèmes fondamentaux des HMMs, Et l'utilisation de ce modèle.
- **Chapitre 4** : Détaille l'approche proposée pour la reconnaissance des postures et des mouvements.
- **Chapitre 5** : Décortique la partie implémentation et expose quelques testes avec les résultats.

Ce travail s'achève par une conclusion générale et quelques perspectives sur les travaux futurs dans ce domaine de recherche.

# Chapitre 1 :

---

---

## La vision par ordinateur

---

---

### 1. Introduction

La vision est le sens qui nous fournit le plus d'informations sur le monde extérieur, elle nous permet de voir, de décrire, et de reconnaître les objets qui sont présent dans la scène.

La vision par ordinateur consiste à reproduire les résultats obtenus par la vision humaine sur un ordinateur en utilisant des moyens informatiques en remplaçant l'œil par une caméra et le cerveau par un ordinateur. Elle est une discipline dont les premières bases théoriques ont été jetées dans les années 60. Depuis, étant donné le spectre très large d'applications industrielles, militaires, aérospatiales et médicales qui peut être envisagé, la vision par ordinateur a vite fait de dépasser le cadre relativement restreint des laboratoires de recherches.

Ce chapitre donne une vue globale sur la vision par ordinateur.

### 2. Définition générale de vision par ordinateur

La vision par ordinateur est la science et la technologie des machines qui peuvent voir, où voir dans ce cas signifie que la machine est capable d'extraire des informations d'une image nécessaire pour résoudre certaines tâches.

En tant que discipline scientifique, la vision par ordinateur est préoccupée par la théorie derrière les systèmes artificiels qui extraient des informations à partir d'images. Les données des images peuvent prendre plusieurs formes, telles que des séquences vidéo, vues de plusieurs caméras, ou des données multidimensionnelles.

En tant que discipline technologique, la vision par ordinateur vise à appliquer ses théories et modèles pour la construction de systèmes de vision par ordinateur. Les exemples d'applications de la vision par ordinateur comprennent des systèmes pour:

- Reconnaissance d'objets dans les images.
- processus de contrôle (par exemple, un robot industriel ou d'un véhicule autonome).
- Détection des événements (par exemple, pour une surveillance visuelle ou de comptage de personnes).

- Organisation de l'information (par exemple, des bases de données d'indexation d'images et de séquences d'images).
- Modélisation des objets ou des environnements (par exemple, l'inspection industrielle, l'analyse d'images médicales ou la modélisation topographique).
- Interaction (par exemple, que l'entrée d'un dispositif pour l'interaction homme-machine).

Les sous-domaines de la vision par ordinateur comprennent la reconstruction scène, la détection d'événements, le suivi vidéo, la reconnaissance d'objets, l'apprentissage, l'indexation, l'estimation de mouvement, et la restauration d'images [1].

### ***3. Les approches en vision par ordinateur***

La recherche en vision par ordinateur se répartit en deux courants de pensée qui, avec le temps, tendent à se rapprocher et à se compléter [2]. On distingue :

- L'approche classique dite vision orientée modèle qui consiste à effectuer la reconstruction 3D d'entités observées, puis à comparer la reconstruction à un modèle 3D pré-acquis de ces entités en utilisant des techniques de rétro-projection et de mise en correspondance ;
- L'approche récente de la vision par apparence qui, comme son nom l'indique, s'appuie sur toutes les manifestations visuelles possibles des entités observées, c'est-à-dire, en théorie, sur toutes les images de ces entités prises de tous les points de vue et éclairages possibles.

Dans la première approche, les modèles se font dans le repère de la scène (ce sont donc des modèles 3D). En vision par apparence, les modèles sont construits dans le repère image (ce sont donc des modèles 2D).

### ***4. Les Outils Fondamentaux de la Vision Par Ordinateur***

Il existe une multitude d'outils de vision par ordinateur, dans ce chapitre on présente ceux qui constituent la base fondamentale en vision par ordinateur.

#### ***4.1. Détection de Contours [3]***

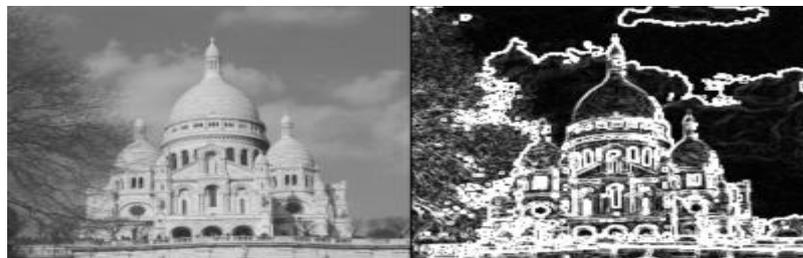
*La* détection de contour est une étape préliminaire à de nombreuses applications de l'analyse d'images. Les contours constituent en effet des indices riches, au même titre que les points d'intérêts, pour toute interprétation ultérieure de l'image.

*Le* principe de la détection de contours repose donc sur l'étude des dérivées de la fonction d'intensité dans l'image : les extrema locaux du gradient de la fonction d'intensité et les passages par zéro du laplacien. La difficulté réside dans la présence de bruit dans les images.

*Le* but de la détection de contours est de repérer les points d'une image numérique qui correspondent à un changement brutal de l'intensité lumineuse. Ces changements de propriétés de l'image traduisent en général des événements importants ou des changements dans les propriétés du monde. Ils incluent des discontinuités dans la profondeur, dans l'orientation d'une surface, dans les propriétés d'un matériau et dans l'éclairage d'une scène. La détection de contour est un champ de la recherche qui appartient au traitement d'image et à la vision par ordinateur, particulièrement dans le domaine de l'extraction de caractéristiques.

*La* détection des contours d'une image réduit de manière significative la quantité de données et élimine les informations qu'on peut juger moins pertinentes, tout en préservant les propriétés structurelles importantes de l'image. Il existe un grand nombre de méthodes de détection des contours de l'image mais la plupart d'entre elles peuvent être regroupées en deux catégories. La première recherche les extremums de la dérivée première, en général les maximums locaux de l'intensité du gradient. La seconde recherche les annulations de la dérivée seconde, en général les annulations du laplacien ou d'une expression différentielle non-linéaire.

La Figure 1.1 illustre d'un exemple de détection de contour de sobel:



**Figure 1.1:** Exemple de détection de contours (Sobel).

Les opérations décrites précédemment se traduisent par des filtres à appliquer à l'image. Canny a cherché à définir des critères afin d'obtenir un filtre optimal pour la détection de contour. Ces critères sont les suivants :

- bonne détection : détecter un maximum de contours ;
- bonne localisation : les points détectés doivent être les plus proches possibles du vrai contour ;
- réponse unique : minimiser le nombre de contours détectés plusieurs fois.

Ces critères se traduisent par des conditions sur la réponse impulsionnelle du filtre et débouchent sur des détecteurs de contours très performants.

#### ***Filtre dérivées premières:***

Le filtre le plus simple consiste à calculer les différences entre pixels voisins sur les horizontales puis sur les verticales. Chaque extremum correspond à un point d'un contour.

#### ***Filtre de Prewit:***

Il introduit un flou, chacune des deux matrices étant le produit du filtre dérivation dans la direction considérée par un filtre de flou rectangulaire selon l'autre direction.

#### ***Filtre de Sobel:***

La technique précédente est améliorée en remplaçant le filtre rectangulaire par un filtre triangulaire.

#### ***Filtre de Canny:***

C'est un filtre de Sobel précédé par un lissage gaussien et suivi par un seuillage. Ce filtre est conçu pour être optimal, au sens de trois critères.

La détection de contour est très utile en traitement d'images, c'est par exemple une étape indispensable à la reconnaissance de formes. Ces algorithmes sont également utilisés en imagerie médicale, cartographie, etc.

## *4.2. Segmentation*

La segmentation d'image est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis. Les pixels sont ainsi regroupés en régions, qui constituent un pavage ou une partition de l'image. Il peut s'agir par exemple de séparer les objets du fond. Si le nombre de classes est égal à deux, elle est appelée aussi binarisation.

Si l'homme sait naturellement séparer des objets dans une image c'est grâce à des connaissances de haut niveau (compréhension des objets et de la scène). Mettre au point des algorithmes de segmentation de haut niveau (chaque région est un objet sémantique) est encore un des thèmes de recherche les plus courants en traitement d'images.

La segmentation est une étape primordiale en traitement d'image. À ce jour, il existe de nombreuses méthodes de segmentation. Nous pouvons distinguer trois types de méthodes de segmentation:

- a) Segmentation fondée sur la classification
- b) Segmentation fondée sur les contours.
- c) Segmentation fondée sur les régions, On y trouve par exemple: la croissance de région et, la décomposition/fusion.

### *4.2.1. Segmentation par classification*

Parmi les techniques de segmentation d'images, la classification est l'une des procédures les plus utilisées. La classification d'une image consiste à partitionner cette image en un ensemble de classes disjointes. A chaque point de l'image on attribue une étiquette parmi les étiquettes qui correspondent chacune à une classe [4].

### *4.2.2. Segmentation par détection de contour*

La recherche des contours dans une image numérique est un des problèmes les plus étudiés depuis l'origine des travaux sur l'imagerie numérique, les approches contours travaillent sur les discontinuités de la fonction d'intensité dans les images afin de déterminer les contours des régions [5].

*Un contour peut être considéré comme un ensemble de pixels formant une frontière entre deux ou plusieurs régions voisines (suivant un critère de similarité donné, par exemple un niveau de gris) [6].*

*La détection d'un contour est généralement basée sur la détection du changement à la frontière de deux régions. De ce fait, on peut pressentir que le bruit dans l'image (anomalies sur les niveaux de gris par exemple) va compliquer ce travail de détection [6].*

*Les méthodes de détection des contours comportent deux grandes étapes [6] :*

- D'abord trouver les pixels censés appartenir à un contour (éventuellement avec une mesure de certitude) en s'appuyant sur une propriété particulière (par exemple un changement local de niveau de gris) en appelant ces points : des points-candidats ou des points contour.
- Puis relier ces points contour de façon à obtenir de véritable contour.

*L'intérêt principal des méthodes de segmentation selon l'approche frontières est de minimiser le nombre d'opérations nécessaires en cas d'itération du processus sur des séries d'images peu différentes les unes des autres (cas des images vidéo notamment). En effet, une fois que les contours des régions ont été trouvés dans la première image, l'application du modèle déformable à l'image suivante est plus efficace que de tout recalculer, si la différence entre les images est peu importante.*

#### *4.2.3. Segmentation par région*

*Dans ce type d'approche, On ne suppose plus comme dans la première famille que l'on peut trouver des contours autour des différents objets que l'on veut mettre en évidence, on cherche plutôt à trouver dans quelle région de l'image ces objets se trouvent. Les régions sont définies comme des ensembles de pixels qui possèdent des propriétés communes. Le choix des propriétés étudiées sont déterminés par le (ou les) critères de segmentation. Deux approches de segmentation région sont largement utilisées à savoir:*

- **Segmentation par croissance de régions** : La croissance de région est une technique contextuelle ; elle prend en considération les attributs locaux des pixels. C'est une technique essentiellement ascendante.

- **Segmentation par division et fusion (Split\_ and\_ merge) :** L'algorithme Split\_and\_merge a été proposé par Horowitz et Pavlidis [7]. Le processus est décomposé en deux étapes. L'image initiale peut être une première partition résultant d'une analyse grossière. Dans la première étape, ou division, on analyse individuellement chaque région  $X_i$ . Si celle-ci ne vérifie pas le critère d'homogénéité, alors on divise cette région en blocs (le plus généralement en 4 quadrants) et l'on réitère le processus sur chaque sous région prise individuellement. On peut tout à fait initier le processus en considérant que la première région est composée de toute l'image. Dans la deuxième étape, ou réunion, on étudie tous les couples de régions voisines ( $X_k, X_i$ ). Si l'union de ces deux régions vérifie le critère d'homogénéité, alors, on fusionne les régions. La principale difficulté de cette approche réside bien sûr dans le parcours de l'ensemble de tous les couples de régions voisines.

### 4.3. Calibrage de Caméra [8]

Le calibrage géométrique d'une caméra consiste à déterminer la relation mathématique existant entre les coordonnées des points 3D de la scène observée et les coordonnées 2D de leur projection dans l'image (points-image). Cette étape de calibrage constitue le point initial pour plusieurs applications de la vision artificielle, comme par exemple la reconnaissance et la localisation d'objets, le contrôle dimensionnel de pièces, la reconstruction de l'environnement pour la navigation d'un robot mobile, etc.

La figure 1.2 illustre l'estimation de transfert en calibrage.



**Figure 1.2 :** Calibrer une caméra consiste à estimer sa fonction de transfert

Le calibrage d'une caméra est particulièrement important lorsque l'on doit obtenir, à partir des images acquises, des informations métriques en vue d'applications de mesures dimensionnelles. Pour obtenir des mesures dimensionnelles précises, il est indispensable de prendre en compte les distorsions géométriques induites par le système optique utilisé.

Calibrer une caméra, c'est choisir un modèle de caméra a priori et déterminer ensuite les paramètres de ce modèle.

*Dans le cas du modèle sténopé (avec ou sans distorsion), il s'agit d'estimer les paramètres intrinsèques de la caméra, et sa position et orientation par rapport au repère du monde qui a été choisi (paramètres extrinsèques).*

*A vrai dire, lorsqu'on calibre une caméra, c'est essentiellement pour déterminer ses paramètres intrinsèques qui, comme leur nom l'indique, sont intrinsèques à la caméra et ne changent pas si l'on déplace la caméra. Des méthodes spécifiques (dites de localisation) ont été développées pour déterminer la position d'une caméra par rapport à un repère de travail lorsqu'on connaît déjà ses paramètres intrinsèques.*

*Généralement, ce problème de calibrage est résolu en utilisant un objet de calibrage spécifique (appelé mire) qui fournit des points 3D connus dans le repère du monde.*

*De nombreuses méthodes de calibrage ont été proposées. Au fil des années, ces méthodes sont devenues de plus en plus sophistiquées pour conduire à un calibrage de plus en plus précis, tout en étant de plus en plus facile à mettre en œuvre.*

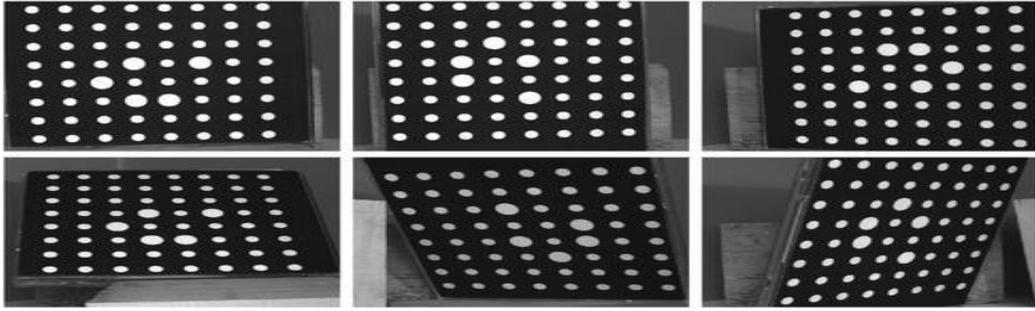
*La méthode la plus performante aujourd'hui, est la méthode consiste à acquérir  $n$  images d'une mire (plane) composée de  $p$  points déplacée librement (rotations et translations) dans le champ de vue de la caméra (Figure 1.3). La méthode de calibrage qui va être décrite est dite de type photo-grammétique. Elle permet d'estimer en même temps tous les paramètres du modèle de caméra ainsi que les points tridimensionnels de la mire. Par conséquent, la géométrie de la mire de calibrage n'a pas besoin d'être connue avec précision a priori.*

*Lors du déplacement de la mire de calibrage, il est important de bien couvrir l'ensemble du champ de vue de la caméra de façon à calibrer correctement la distorsion (qui est généralement plus importante au bord des images qu'au centre).*

*Les points de la mire peuvent être les nœuds d'intersections de droites horizontales et verticales (grille), les coins d'une mire de damiers, ou bien les centres de taches circulaires.*

*Les points de la mire, qui sont extraits par des procédures spécifiques de traitement d'images, fournissent les mesures.*

La Figure 1.3 illustre l'utilisation des images composées de 64 taches circulaires, utilisée pour le calibrage d'une caméra.



*Figure 1.3 : Exemple d'une séquence de 6 images d'une mire de calibrage.*

#### **4.4. Reconnaissance d'objet**

*L*a reconnaissance d'objets est un domaine très actif en vision par ordinateur. L'approche générale consiste à considérer qu'on dispose d'une base de données où sont stockés des modèles d'objets et une vue du monde réel. On doit répondre à une question essentielle : L'objet extrait, existe-t-il dans la base de données ? Le problème dépend de la modélisation d'objets adoptée. Dans la majorité des cas cette modélisation est purement géométrique : Un ensemble de caractéristiques appelées primitives. L'opération de reconnaissance consiste en l'appariement de ces caractéristiques et celles décelées dans l'image.

*O*n utilise des techniques d'alignement qui consistent à transformer le repère des modèles dans celui des données (image). Des travaux significatifs dans ce sens ont été présentés par Ayache et Faugera [9] (système HYPER), Lowie [10], (système SCERPO) etc.

*E*n réalité, la reconnaissance doit prendre en compte d'autres caractéristiques telles que la couleur, la texture, le non rigidité, etc. Des travaux importants [11], système SEEMORE ont été avancés en s'appuyant sur des modèles de réseaux de neurones pour stocker de l'information dans une phase d'apprentissage et d'activer cette information dans une autre phase : la reconnaissance. Ces approches travaillent directement dans l'espace d'image et non pas dans une représentation structurelle d'objets.

#### **4.5. Reconstruction 3D**

*L*a reconstruction 3D de surfaces à partir d'une séquence d'images est un sujet de recherche très actif dans la communauté de la vision par ordinateur. Il existe de nombreuses méthodes de reconstruction 3D. Les approches existantes diffèrent beaucoup selon les techniques utilisées ainsi que par les connaissances a priori du système et de l'environnement. On peut citer les méthodes pour modéliser la végétation comme les arbres par exemple [12],

les approches de numérisation 3D de visage par stéréoscopie active, ou la modélisation 3D de bâtiments pour la création de bases de données urbaines 3D.

*On* peut répertorier une grande partie des méthodes de reconstruction dans ces cinq Catégories:

- Reconstruction à partir de la géométrie épi-polaire ;
- Reconstruction à partir de la texture ;
- Reconstruction à partir des silhouettes ;
- Reconstruction à partir du mouvement de la camera.

## ***5. Quelques applications concrètes***

*La* vision artificielle trouve des applications dans des domaines très variés. Voici un petit panel de domaines où l'on peut la rencontrer.

### ***5.1. La surveillance intelligente***

*L'un* des domaines les plus grands où s'applique la vision artificielle est la vidéosurveillance.

*Dans* ce domaine, la vision par ordinateur est utilisée pour suivre des gens dans des lieux fréquentés, de manière à les compter, ou à détecter des comportements particuliers (une agression, une personne qui tombe, un attroupement...). Une autre application courante est la surveillance du trafic routier, notamment pour compter les voitures qui passent, détecter des embouteillages, des accidents... ou des excès de vitesse...

*Dans* ce cas précis, la vision par ordinateur sert avant tout à faciliter le travail des employés qui sont attelés à la vidéosurveillance. Il s'agit bien souvent de contrôle, de suivi, et de prises de décision.

## *5.2. L'assistance et les interfaces homme-machine*

*D*ans un tout autre domaine, on peut rencontrer la vision dans des applications d'aide aux personnes handicapées (ou non), de manière à leur permettre d'utiliser leur ordinateur. On peut retrouver aussi des interfaces sophistiquées basées sur le geste.

## *5.3. Les jeux vidéo*

*L'*exemple le plus populaire est celui de la console Wii de Nintendo. La technologie qui permet de capter les mouvements de la manette (la WiiMote) fait partie du cadre de la vision par ordinateur (bien qu'il s'agisse avant tout de détection par infra-rouges). Mais quitte à se baser sur des informations purement visuelles, on peut aussi mentionner le fameux projet Natal de Microsoft, dont le but est de s'affranchir une bonne fois pour toutes de la manette de jeu.

*C*ette liste est très loin d'être exhaustive. La vision par ordinateur peut trouver des applications dans une quasi-infinité d'autres domaines que ceux-ci, comme l'aide au diagnostic médical, la robotique, la surveillance de chaînes de production dans l'industrie.

Toutes ces applications reposent sur le même type de traitements.

## *6. Conclusion*

*La* Vision par ordinateur est devenue un domaine de recherche très actif. Il est fortement stimulé par de nombreuses applications prometteuses telles que la surveillance intelligente, la réalité virtuelle, l'interface utilisateur avancée, etc.

*D*ans ce chapitre on a traité la vision par ordinateur donnant une définition générale, puis les approches de la vision par ordinateur. Après, on a mentionné quelques outils fondamentaux de la vision tel que la détection des contours, segmentation, le calibrage, la reconnaissance d'objet et la reconstruction 3D, ensuite quelques applications et des exemples concrets. Parmi les domaines d'application de la vision par ordinateur, on traite l'analyse de mouvement humain dans le chapitre suivant.

## *Chapitre 2:*

---

---

### *Postures et mouvement humain*

---

---

#### *1. Introduction*

Comme l'un des domaines de recherche les plus actifs en vision par ordinateur, l'analyse de mouvements humains a pour objectifs de détecter, suivre et identifier les personnes, et plus généralement, pour interpréter les comportements humains, à partir de séquences d'images avec des personnes. L'analyse du mouvement de l'homme a suscité de grands intérêts dans la communauté de chercheurs de vision par ordinateur en raison de ses applications prometteuses dans de nombreux domaines, tels que la surveillance, l'interface utilisateur, le stockage d'images basées sur le contenu et l'extraction, la vidéo conférence, l'analyse des performances athlétiques, la réalité virtuelle, etc...

L'analyse du mouvement humain fait intervenir plusieurs processus de traitement d'images tels que la segmentation d'objets en mouvement, le suivi temporel, la détection de peau, les modèles de corps humain et la reconnaissance d'actions ou de postures.

Ces dernières années, l'analyse du mouvement de l'homme a été présentée dans un certain nombre de grandes revues internationales telles que IJCV (International Journal of Computer Vision), CVIU (Computer Vision and Image Understanding), PAMI (IEEE Transactions on Pattern Recognition and Machine Intelligence) et IVC (Image and Vision Computing), ainsi que de prestigieuses conférences et ateliers internationaux tels que ICCV (Conférence internationale sur la vision par ordinateur) et CVPR (IEEE Conférence internationale sur la vision par ordinateur et reconnaissance de formes), [13].

#### *2. Grands domaines d'application*

##### *2.1. Surveillance Vidéo*

Le besoin impérieux des systèmes de surveillance intelligents, écoule de ces zones de sécurité sensibles comme les banques, les grands magasins, les parcs de stationnement, et les frontières [14]. Les caméras de surveillance sont déjà répandues dans les établissements commerciaux, tandis que les sorties camera sont habituellement enregistrés en bandes ou stockés dans les archives vidéo.

## *2.2. Interface utilisateur Avancée*

*Un* autre domaine d'application important est les interfaces utilisateur avancées dans lesquels l'analyse du mouvement humain est généralement utilisée pour assurer le contrôle et le commandement. De manière générale, la communication entre les gens est principalement réalisée par la parole. Par conséquent, la compréhension de la parole a déjà été largement utilisée dans les interfaces homme-machine au début. Cependant, elle est soumise aux restrictions de bruit dans l'environnement et de la distance. La vision est très utile pour compléter la reconnaissance vocale et la compréhension du langage naturel pour une communication plus naturelle et intelligente entre l'homme et la machine. C'est-à-dire, des indices plus détaillées peuvent être obtenues par des gestes, les poses du corps, les expressions faciales, etc. [15].

## *2.3. L'identification et diagnostic basé sur le mouvement*

*Il* est particulièrement utile pour segmenter différentes parties du corps humain dans une image, suivre le mouvement des joints sur une séquence d'images, et de récupérer la structure sous-jacente du corps 3-D pour l'analyse des performances sportives (la formation). Avec le développement des bibliothèques numériques, l'interprétation de séquences vidéo automatiquement en utilisant l'indexation basée sur le contenu permettra d'économiser d'énormes efforts de l'homme dans le tri et la récupération des images ou vidéo dans une énorme base de données. L'analyse de la marche traditionnelle [16] [17] vise à fournir un diagnostic médical et un soutien aux traitements, tandis que la démarche de l'homme peut aussi être utilisée comme une caractéristique biométrique pour l'identification des personnes [18] [19].

*Certaines* autres applications de l'analyse du mouvement basée sur la vision se trouvent dans les systèmes de formation personnalisée pour différents sports, diagnostics médicaux des patients orthopédiques, chorégraphie de danse et de ballet, etc...

*En* outre, l'analyse du mouvement humain montre son importance dans d'autres domaines connexes. Par exemple, les applications typiques incluent la réalité virtuelle, les jeux, studios virtuels, l'animation de personnages, téléconférence, etc.

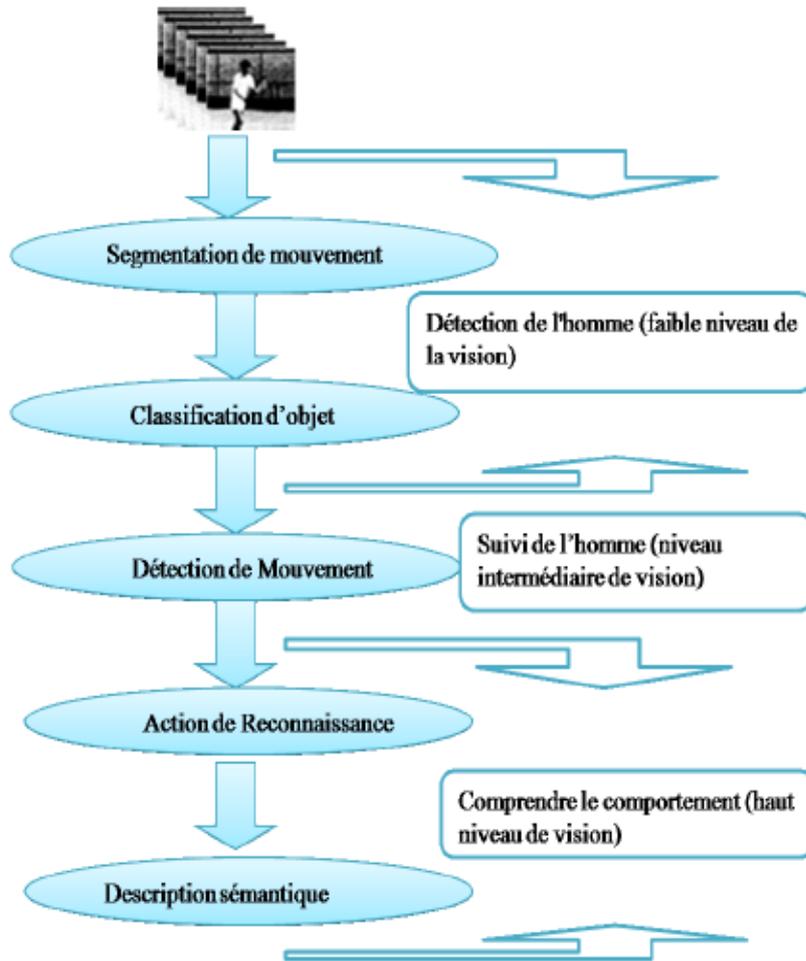
*En* ce qui concerne les jeux d'ordinateur.[20] sont concernés, ils ont été très répandue dans le divertissement. Peut-être que les gens sont surpris de le réalisme des humains virtuels et les actions simulées dans les jeux d'ordinateur. En fait, cette bénéficie grandement de

l'infographie portant sur l'élaboration des modèles réalistes du corps humain et de la synthèse du mouvement humain fondé sur la connaissance de l'acquisition de modèle du corps humain, la récupération du pose de corps, l'analyse du comportement de l'homme, etc. En outre, il est évident que l'image fondée sur un modèle de codage (par exemple, codant que la pose de la face suivi dans les images de façon plus détaillée que le fond indifférent dans un cadre vidéophone) apportera très faible débit compression vidéo pour le stockage des images plus efficace et la transmission.

### ***3. Processus d'analyse du mouvement humain***

L'intérêt croissant pour l'analyse du mouvement humain a permis des progrès significatifs ces dernières années, en particulier sur les questions de vision de haut niveau tel que l'activité humaine et la compréhension du comportement. Nous nous concentrerons sur un aperçu général sur le processus global d'un système d'analyse du mouvement humain. Trois tâches principales dans ce processus (détection de l'homme, le suivi de l'homme et la reconnaissance du mouvement humain) seront particulièrement revues. La majorité des travaux antérieurs dans l'analyse du mouvement humain sont réalisées dans le cadre du suivi et de la reconnaissance d'action.

La figure 2.1 [13] illustre le processus général d'analyse de mouvement humain.



*Figure 2.1 : Un cadre général d'analyse du mouvement humain.*

## 4. La Détection

Presque tous les systèmes d'analyse du mouvement humain commencent par la détection de l'homme. La détection vise à segmenter les zones correspondant à des personnes du reste de l'image. C'est une question importante dans un système d'analyse de mouvement humain puisque les processus subséquents telles que le suivi et la reconnaissance d'action sont grandement dépendants de la qualité de la détection. Ce processus implique généralement la segmentation de mouvement.

### 4.1. Segmentation de mouvement

La segmentation de mouvement dans des séquences vidéo est connue pour être un problème important et difficile qui vise à détecter les régions correspondant aux objets

mobiles tels que les véhicules et les personnes dans les scènes naturelles. La détection des blobs mobile fournit le centre d'attention pour les processus ultérieurs tels que le suivi et l'analyse d'activité parce que seuls les pixels changeant doivent être considérées. Cependant, les changements de la météo, l'éclairage, les ombres et les mouvements répétitifs d'un fond encombré rendent la segmentation de mouvement difficile à traiter rapidement et de manière fiable.

À l'heure actuelle, la plupart des méthodes de segmentation utilisent les informations spatiales ou temporelles des images. Plusieurs approches classiques de segmentation de mouvement sont décrites ci-après.

#### *4.1.1. Soustraction de fond*

La soustraction de fond [21] est une méthode particulièrement populaire pour la segmentation de mouvement, en particulier dans des situations avec un fond relativement statique. Elle tente de détecter les zones en mouvement dans une image par différence entre l'image courante et une image de fond de référence dans un mode pixel par pixel. Toutefois, cette technique est extrêmement sensible aux dynamiques de changements de scènes en raison d'événements extérieurs et d'éclairage.

Plusieurs recherches se sont intéressées à la construction de différents modèles d'arrière-plan adaptatif afin de réduire l'influence des changements dynamiques de scène sur la segmentation du mouvement. Par exemple, certaines études préliminaires données par Karman et Brandt et Kilger [22], on proposé un modèle d'arrière-plan adaptatif basé sur le filtrage de Kalman pour s'adapter aux changements temporels de la météo et de l'éclairage.

#### *4.1.2. Méthodes statistiques*

Récemment, certaines méthodes statistiques pour extraire les régions de changements de l'arrière-plan se sont inspirées des méthodes de base de soustraction de fond (modèle d'arrière-plan adaptatif). Les approches statistiques utilisent les caractéristiques des pixels individuels ou de groupes de pixels pour construire des modèles d'arrière-plan plus avancés, les statistiques des arrière-plans peuvent être mises à jour dynamiquement au cours du traitement. Chaque pixel de l'image en cours peut être classé en premier plan ou d'arrière-plan en comparant les statistiques du modèle de base actuel. Cette approche est de plus en plus populaire en raison de sa robustesse au bruit, à l'ombre, au changement des conditions d'éclairage, etc.

Une étude récente de Haritaoglu et al. [23] présente la construction du modèle statistique adaptatif par un représentant de chaque pixel avec trois valeurs: les valeurs d'intensité minimale et maximale, et la différence d'intensité maximale entre les images consécutives observées au cours de la période de formation. Les paramètres du modèle sont mis à jour périodiquement.

#### *4.1.3. Différence inter-images*

Le problème peut se définir simplement de la manière suivante : à partir d'une séquence d'images capturée par une caméra fixe, nous voulons détecter les objets ne faisant pas partie de la scène. La solution peut se décrire naïvement : la détection s'effectue par soustraction d'une image statique de la scène à l'image courante.

La différence d'images est une technique relativement rapide utilisée pour récupérer les objets en mouvement dans une scène. Cette méthode est étudiée depuis la fin des années 1970 [24]. Les images consécutives sont soustraites les unes aux autres, l'application d'un seuil sur l'image des différences permet ensuite de récupérer les pixels correspondant à des régions en mouvement.

Plusieurs problèmes immédiats viennent à l'esprit lors de l'utilisation de cette technique. Si une région en mouvement et une région de la scène possèdent une même couleur et se superposent, elles seront supprimées. Les pixels subissant des variations brutales de luminosité seront détectés comme faisant partie de l'objet en mouvement. Un objet mobile que l'on cherche à détecter peut rester immobile pendant une durée  $\delta t$  et ne sera plus détecté. Enfin les objets du fonds "désoccultés" sont interprétés comme en mouvement.

#### *4.1.4. Flot optique*

L'utilisation du flux optique procure une information de mouvement pour chaque pixel de l'image. Ainsi, il mesure les vecteurs de déplacement à partir de l'intensité des pixels de deux images consécutives ou temporellement rapprochées. Dans un contexte de détection de mouvement, les pixels inactifs posséderont alors une vitesse nulle contrairement aux pixels appartenant à des objets dynamiques. Une classification sous forme de regroupement est donc nécessaire afin d'isoler et de localiser les zones représentant du mouvement. Cette technique a notamment été utilisée pour la détection de piétons [25].

L'inconvénient majeur de l'utilisation du flux optique est la somme importante de calculs à réaliser pour l'estimation du mouvement. De plus, si certaines parties d'un objet ne sont pas en mouvement, elles seront complètement ignorées par cette méthode. Ce pourrait être le cas par exemple d'une séquence vidéo contenant une personne assise par terre et agitant les bras. Dans cette situation bien précise, le corps de la personne ne serait pas détecté contrairement à ses bras [26].

## 5. Le Suivi

Le suivi d'objets dans des flux vidéo a été un sujet populaire dans le domaine de la vision par ordinateur. Le suivi est un enjeu particulièrement important dans l'analyse du mouvement humain, car il sert comme un moyen de préparer les données pour l'estimation de la pose et la reconnaissance d'action. Contrairement à la détection de l'homme, le suivi de l'homme appartient à un problème de vision par ordinateur de niveau supérieur. Toutefois, les algorithmes de suivi dans l'analyse du mouvement humain ont, en général, une intersection importante avec une segmentation de mouvement au cours du traitement.

Typiquement, le suivi au fil du temps implique des objets correspondant à des images consécutives à l'aide des caractéristiques telles que des points, des lignes ou des blobs. C'est-à-dire, le suivi peut être considéré comme équivalent à l'établissement de relations cohérentes des caractéristiques de l'image entre les trames concernant la position, la vitesse, la forme, la texture, la couleur, etc.

Le suivi peut être divisé en différentes catégories selon différents critères. Des classements possibles peuvent être:

- Suivi des parties du corps humain tels que les mains, le visage, et de la jambe et le suivi de tout le corps [27].
- Nombre de vues considérées: un seul point de vue [28], multi-vues [29],
- le suivi par vue omnidirectionnelle [30].

Les techniques de suivi peuvent aussi être regroupées selon d'autres critères tels que la dimension de l'espace de suivi (2D vs 3D), l'environnement du suivi (à l'intérieur à l'extérieur), le nombre d'humain suivis (un seul humain, humains multiples, des groupes humains), l'état de la caméra (mouvement vs stationnaire), multiplicité de capteur (monoculaire vs stéréo), etc.

Différentes méthodes sont utilisées dans le processus de suivi. Le suivi peut être à base de modèles, de régions, ou de caractéristiques. Ces méthodes sont détaillées ci-dessous.

### *5.1. Suivi à base de modèle*

Pour détecter et identifier les différents membres du corps, il faut avoir un modèle géométrique en 2D (avec ou sans modèle explicite de la forme) ou en 3D. Le suivi avec modèle explicite compare les données issues de l'image avec un modèle de l'objet ou de la personne à suivre. Cette méthode requiert le développement d'un modèle 2D ou 3D de la personne selon l'application.

- Si le modèle du corps est simple, il sera facile à implémenter, rapide, mais sujet aux occultations et peu précis par rapport aux variations de posture, d'angle, et d'apparence (fonction du point de vue et des occultations). Pour les applications où la capture de la pose n'est pas nécessairement exacte comme le suivi de personnes pour la télésurveillance, l'approche 2D est appropriée. Il en est de même des applications avec une seule personne impliquant des contraintes sur le mouvement et un point de vue simple (estimation de la posture de la main en reconnaissance de la langue des signes face à la caméra, reconnaissance de la marche latéralement à la caméra);
- Les approches 3D correspondent aux applications de suivi de mouvements complexes et non contraints (interactions entre personnes comme se serrer la main, danser ou se battre). Les approches 3D sont plus exactes et résolvent les occultations et collisions, en revanche elles ne sont pas adaptées au temps réel.

Les méthodes basées modèles sont robustes aux occultations car elles possèdent une connaissance a priori d'un modèle de la forme, contrairement aux autres méthodes sans modèle, mais elles demandent un coût de calcul important. L'information structurelle du modèle de la forme sert à mettre en correspondance les données image avec le modèle, soit par une approche ascendante associant des hypothèses images, soit par une approche descendante où on cherche le modèle ayant le maximum de corrélation avec les données image. Mais elles demandent un coût de calcul important [31].

### *5.2. Suivi basé régions*

L'approche par région se caractérise par l'extraction dans l'image courante de régions dénommées « blobs », ensemble de pixels connexes et regroupés en fonction d'un critère

déterminé, par exemple les pixels dont la valeur est différente avec ceux de l'image précédente, et le suivi des régions homogènes au cours de la séquence. Cette méthode est basée sur la variation du mouvement dans les régions de l'image. Elle ne résout pas les occultations entre objets. L'hypothèse est faite qu'à l'intérieur d'une région, l'apparence est invariante et le mouvement est homogène, par exemple le suivi de « blobs » par filtrage de Kalman [32].

[33] et [34] utilisent la différence absolue entre l'image courante  $I_t$  et une image de référence  $I_o$  :  $I_{résultat} = |I_o - I_t|$ . L'inconvénient de cette méthode est qu'il n'y a pas de mise à jour de l'image de référence  $I_o$ . L'autre méthode très usitée est la différence d'images successives [35] :  $I_{résultat} = \text{Max} (|I_t - I_{t-1}|, |I_{t-1} - I_t|)$ . Elle est utile lorsqu'on ne dispose pas d'image de référence, mais elle ne prend pas en compte les mouvements des régions uniformément colorées. Seules les régions texturées sont détectées [36].

*Les approches de suivi basées régions fonctionnent raisonnablement bien. Toutefois, des difficultés apparaissent dans deux situations importantes. La première est celle de longues ombres, il peut en résulter le raccordement de blobs qui auraient dû être associés à des personnes distinctes. Ce problème peut être résolu dans une certaine mesure en faisant usage de la couleur ou en exploitant le fait que les régions d'ombre ont tendance à être dépourvue de texture. Le problème le plus grave, et jusqu'ici insolubles, pour le suivi vidéo a été celui des situations d'encombrement. Dans ces conditions, les personnes sont partiellement obstruées les uns les autres au lieu d'être spatialement isolés. Cela rend la tâche de segmentation individuelle de chaque personne très difficile. La résolution de ce problème peut exiger des systèmes de suivi à l'aide de plusieurs caméras.*

### *5.3. Suivi basé caractéristiques*

*Dans les méthodes de suivi de caractéristiques, l'objectif est de détecter des descripteurs liés à des points particuliers, et décrivant l'objet par un ensemble d'attributs géométriques (points, segments, courbes paramétriques, arêtes, contours), ou des régions de l'image. Ces méthodes ont l'avantage d'une bonne robustesse aux occultations car des associations qui n'ont pas pu se faire sur certains points de l'objet cachés dans l'image, ne mettent pas en échec le suivi sur l'ensemble des points.*

## *6. Estimation de pose*

L'estimation de pose désigne le processus d'estimation de la configuration de la structure sous-jacente de la cinématique ou de l'articulation squelettique d'une personne. Ce processus peut être une partie intégrante du processus de suivi comme dans les approches d'analyse par synthèse basée sur un modèle, ou peut être effectué directement à partir des observations sur une base per-frame (image par image). L'étude [37] a classé les algorithmes d'estimation de pose en trois catégories fondées sur leur utilisation d'un modèle a priori de l'homme:

### *6.1. Sans modèle*

Cette classe couvre les méthodes où il n'existe pas de modèle explicite a priori. Une tendance récente pour surmonter les limitations de suivi sur de longues séquences a été l'investigation de détection directe de pose sur les frames. Deux approches qui entrent dans cette catégorie d'estimation de pose sans modèle ont été étudiées: assemblages probabilistes de parties où les parties du corps individuels sont d'abord détecté et ensuite assemblés pour estimer la pose 2D; méthodes basées sur l'exemple qui apprend directement le mapping de l'espace d'image 2D à l'espace modèle 3D.

#### *6.1.1. Assemblage probabiliste de parties*

L'assemblage probabiliste des parties a été introduit pour une estimation directe ascendante de pose 2D en commençant par la détection des positions probables des parties du corps humain, puis l'assemblage de ces positions pour obtenir la configuration qui correspond le mieux aux observations. Un avantage potentiel pour la détection et suivi est que la pose peut être estimée indépendamment à chaque image, ce qui permet l'estimation de pose pour les mouvements rapides.

#### *6.1.2. Méthodes basées exemples*

Un certain nombre de méthodes basées sur l'exemple pour l'estimation de pose humaine ont été proposé, elles comparent l'image observée avec une base de données d'échantillons. Brand [38] a utilisé un modèle de Markov caché (HMM) pour représenter le mapping entre des séquences de silhouette 2D dans l'espace d'image et le mouvement du squelette dans un espace de poses 3D. Dans ce travail, le mapping des séquences de mouvements spécifiques a été appris en utilisant des images de rendu de silhouette d'un modèle humanoïde. Le HMM a été utilisée pour estimer la séquence de pose 3D les plus probables à partir d'une séquence de silhouette 2D observée pour un point de vue spécifique.

Les approches basées exemples représentent le mapping entre l'image et l'espace de pose fournissant un mécanisme puissant pour l'estimation directe de pose 3D. Généralement ces approches exploitent un rendu des données de capture de mouvement pour fournir des exemples d'apprentissage avec des poses 3D connues. Une limitation courante des approches basées exemples est la restriction à la pose ou mouvements utilisés dans l'apprentissage. L'extension à un plus large vocabulaire des mouvements peut introduire des ambiguïtés dans le mapping.

## *6.2. Utilisation indirecte de modèles*

Dans cette classe, les méthodes utilisent un modèle a priori dans l'estimation de pose comme une référence ou table de consultation pour guider l'interprétation des données mesurées.

Un certain nombre de chercheurs ont étudié la reconstruction directe de la forme et des mouvements du modèle à partir d'une coque visuelle (visual hull) [39] sans modèle a priori. Mikic et coll. [40] présente un système intégré de récupération automatique d'un modèle du corps humain et du mouvement à partir de plusieurs vues d'une séquence d'images. L'acquisition de modèle est basée sur une approche hiérarchique basée sur des règles de localisation et d'étiquetage de parties de corps. La connaissance préalable de la forme des parties du corps, la taille relative, et la configuration sont utilisées pour segmenter la coque visuelle. Un filtre de KALMAN étendu est ensuite utilisé pour la reconstruction du mouvement humain entre les images.

## *6.3. Utilisation directe de modèles*

Cette classe utilise explicitement une représentation géométrique 3D de forme humaine et une structure cinématique pour reconstituer la pose. La majorité des approches utilisent une méthode d'analyse par synthèse pour optimiser la similarité entre le mode de projection et l'image observée.

## *7. Reconnaissance*

Le champ de la représentation d'action et d'activité et la reconnaissance est relativement ancienne, mais encore immature. Cette zone est actuellement soumise à intenses recherches qui se traduisent également par le grand nombre d'idées et des approches différentes. Les

approches dépendront de l'objectif du chercheur et les objectifs de la reconnaissance d'activités : la surveillance, médicales et de rééducation, de la robotique, l'indexation vidéo et d'animation pour le cinéma et les jeux. Par exemple, dans l'interprétation de la scène, la connaissance est souvent représentée statistiquement et vise à distinguer les activités " régulières" et "irrégulières " .

*D'*autre part, certaines applications de surveillance se concentrent explicitement sur les activités humaines et les interactions entre les humains. Ici, on trouve à la fois des approches holistiques, qui tiennent compte de l'ensemble du corps humain sans tenir compte des parties du corps particulier, et les approches locales. La plupart des approches holistiques tentent d'identifier les "informations" holistique comme le sexe, l'identité, ou de simples actions comme la marche ou la course. Les chercheurs qui utilisent des approches locales sont intéressé par des actions plus subtiles ou tentent de modéliser des actions par la recherche de primitives avec laquelle les actions complexes peuvent être modélisées.

### *7.1. Interprétation de scènes*

*De* nombreuses approches examinent la vue de la caméra dans son ensemble pour essayer d'apprendre et de reconnaître les activités. Ces approches observent le mouvement des objets sans nécessairement connaître leur identité. Ce qui est raisonnable dans des situations où les objets sont suffisamment petits pour être représentés comme des points sur un plan 2D.

*Stauffer* et al. [41] présente un système d'interprétation de scène qui permet de détecter des situations inhabituelles. Le système extrait des caractéristiques telles que la position 2D, la vitesse, la taille et les silhouettes binaires.

### *7.2. Approches holistiques de reconnaissance*

*La* reconnaissance de l'identité d'un être humain, sur la base de sa structure et de la dynamique globale du corps est discutée dans de nombreuses publications. Un intérêt particulier a été porté sur la reconnaissance d'identité par la démarche humaine. D'autres approches utilisant la structure et la dynamique globale du corps sont concernées par la reconnaissance de gestes simples tels que la course et la marche. Presque toutes les méthodes sont à base de silhouette ou de contours [42]. Les techniques sont la plupart du temps holistiques, par exemple, la silhouette entière ou le contour sont pris en compte sans détection des différentes parties du corps.

### *7.3. Reconnaissance basée sur les parties du corps*

De nombreux auteurs sont concernés par la reconnaissance des actions basées sur la dynamique et les paramètres des différentes parties du corps. Certaines approches, par exemple [43], commencent avec des silhouettes et détectent les parties du corps en utilisant une méthode inspirée par le système de W4-system [44]. D'autres utilisent un tracking basé sur un modèle 3D du corps où la reconnaissance de l'action (souvent périodique) est utilisée comme boucle arrière pour appuyer l'estimation de pose.

Dans un travail lié aux [45] Wang et coll. présentent une approche où les contours sont extraits et un contour moyen est calculé pour représenter l'information contour statique. L'information dynamique est extraite en utilisant un modèle détaillé composé de 14 parties rigides du corps, chacune représentée par un cône tronqué. Le filtrage particulaire est utilisé pour calculer la probabilité d'une pose étant donné une image d'entrée. Pour le classement, un classificateur du plus proche voisin (NN) a été utilisé.

## *8. Difficultés de la recherche et orientations*

Même si une grande quantité de travail a été faite dans l'analyse du mouvement humain, de nombreuses questions sont encore ouvertes et méritent des recherches plus poussées, en particulier dans les domaines suivants :

### *8.1. La Segmentation*

La segmentation de mouvement précis est un problème important mais difficile. Les images capturées dans des environnements dynamiques sont souvent influencées par de nombreux facteurs comme les conditions météorologiques, l'éclairage, l'encombrement, l'ombre, l'occlusion, et même le mouvement de la caméra.

### *8.2. La manipulation d'occlusion*

À l'heure actuelle, la majorité des systèmes d'analyse du mouvement de l'homme ne peuvent résoudre efficacement les problèmes de l'auto-occlusion du corps humain et des occlusions mutuelles entre les objets, en particulier la détection et le suivi de plusieurs personnes dans des conditions congestionnées.

### ***8.3. L'utilisation de plusieurs caméras***

*Il* est évident que les futurs systèmes d'analyse du mouvement humain bénéficieront grandement de l'utilisation de plusieurs caméras. La disponibilité d'informations provenant de plusieurs caméras peut être extrêmement utile, car l'utilisation de plusieurs caméras non seulement augmente la zone de surveillance, mais fournit également des points de vue multiples pour résoudre les occlusions efficacement. Le Suivi avec une seule caméra génère facilement l'ambiguïté, ce qui peut être éliminé par une autre vue.

*Pour* les systèmes de suivi multi-caméras, il faut décider quel caméra ou image utiliser à chaque instant du temps. Autrement dit, la coordination et la fusion de l'information entre les caméras sont un problème important.

## ***9. Conclusion***

*L'*analyse du mouvement humain est l'un des domaines de recherche les plus actifs en vision par ordinateur, elle attire de grands intérêts des chercheurs en raison de ses applications prometteuses dans de nombreux domaines. Nous avons présenté un aperçu sur les tâches principales d'un système d'analyse du mouvement humain qui sont : la détection, le suivi, l'estimation de pose et la reconnaissance de posture.

*Même* si une grande quantité de travail a été fait dans ce domaine, de nombreuses questions restent ouvertes telles que la segmentation, la modélisation et la manipulation d'occlusion.

## Chapitre 3 :

---

### Les Chaines de Markov Caché

---

#### 1. Introduction

La modélisation stochastique permet l'utilisation des modèles probabilistes pour traiter les problèmes à information incertaine ou incomplète. Ainsi, les modèles de Markov connaissent un regain d'intérêt tant dans leurs aspects théoriques qu'appliqués.

La théorie des chaînes de Markov est née en 1913, une première application a été développée par Markov pour analyser le langage. Ces travaux ont été utilisés régulièrement mais les premières applications exploitables furent réalisées dans les années 60, telles que les modèles probabilistes d'urnes par Neuwirth, le calcul direct du maximum de vraisemblance ou l'observation de la suite d'états dans une chaîne de Markov. Ceci a permis à la communauté scientifique d'exploiter pleinement le potentiel de ces modèles. C'est dans les années 70 que des chercheurs ont apporté des algorithmes puissants permettant de résoudre les problèmes de reconnaissance, d'analyse et d'apprentissage.

Depuis 1975, les HMM sont utilisés dans de nombreuses applications, principalement dans le domaine de la parole. Ces applications ne se contentent pas de s'appuyer sur la théorie des Modèles de Markov Cachés, mais développent plusieurs extensions théoriques dans le but d'améliorer les modèles. C'est ce qui en a fait leur succès.

Sans prétendre à une présentation exhaustive des modèles de Markov cachés, l'objectif de ce chapitre est de dresser un portrait général de cet outil et de son utilisation.

#### 2. Théorie des chaînes de Markov

Un processus stochastique est un phénomène où intervient le hasard, nous définissons  $X(t)$  une variable aléatoire évoluant en fonction du temps.

##### Exemple

Une suite de lancers de dés 1, 6, 2, 5 d'où  $X_0 = 1, X_1 = 6, X_2 = 2, X_3 = 5$ .

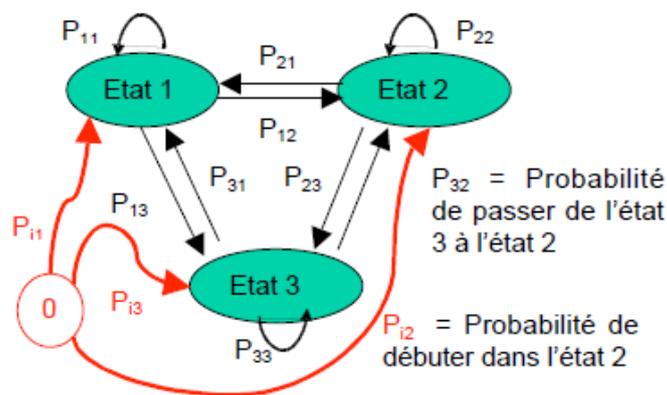
Ce processus est dit markovien si son évolution ne dépend pas de son passé, mais uniquement de son état présent. (Ceci est appelé la propriété de Markov.).

Un processus markovien peut être modélisé par un modèle théorique dit « Modèle de Markov ».

Il existe 2 types de Modèle : **Observable** et **Caché**.

### 2.1. Chaîne Observable

L'évolution du processus de Markov peut être représentée par un graphe de transitions d'états (Figure 3.1) qui fait apparaître la structure du processus selon les règles suivantes:



**Figure 3.1** : Graphe d'un Modèle de Markov Observable

- Les états sont représentés par des sommets. (Etat n) On parle d'alphabet des états :  $S = \{s_1, s_2, \dots, s_n\}$  les états de la chaîne de Markov.
- Les transitions (possibilité de passer d'un état à un autre) sont représentées par des arêtes, elles sont pondérées par leur probabilité. (Flèche)
- Les probabilités de départ : ce sont les probabilités de débiter dans un état ou un autre (point 0) Elles sont regroupées dans un vecteur d'initialisation :  $\Pi = \{\pi_i = P(s_i)\}$  on notera :

$$\sum_{i=1}^N \pi_i = 1$$

Un modèle  $\lambda$  est dit observable car les états sont directement observables, il est caractérisé par une matrice de transition  $A$  et un vecteur d'initialisation  $\pi$ , on note :

$$\lambda = \{\Pi, A\}$$

## 2.2. Chaîne Cachée

Dans un Modèle de Markov Caché les états  $S = \{S_1, S_2, \dots, S_m\}$  sont non observables cependant ils émettent des signaux observables  $O = (O_1, O_2, \dots, O_k)$  qui sont pondérés par leur probabilité. Le modèle  $\lambda$  peut être représenté graphiquement (Figure 3.2), avec :

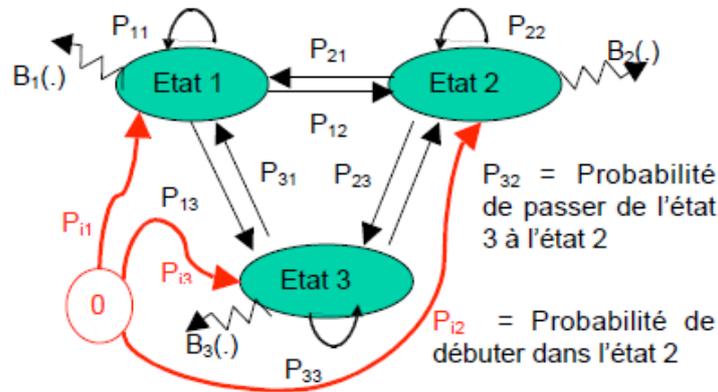


Figure 3.2 : Graphe d'un Modèle de Markov Caché.

Un Modèle de Markov Caché  $\lambda$  est caractérisé par une matrice de transition  $A$ , une matrice d'observation  $B$  et un vecteur d'initialisation  $\pi$ , on note :

$$\lambda = \{\pi, A, B\}$$

### 2.2.1. Définition

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. À chaque transition est associé un symbole observable. Ce symbole est généré à chaque fois que la transition est empruntée.

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 observables. Les modèles de Markov cachés sont utilisés pour modéliser des séquences d'observations. Ces observations peuvent être de nature **discrète** ou **continue**. Si l'ensemble des symboles d'observations  $V$  est fini alors on parle de HMM discret (si  $V$  infini alors HMM continu).

Un HMM peut donc être défini par un quintuplet  $(S, \Sigma, \pi, T, E)$ , ou :

- $S$  est un ensemble de  $N$  états,
- $\Sigma$  un ensemble de  $M$  symboles observables,
- $T = S \times S \rightarrow [0, 1]$  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  $s_0$ ,
- $E = S \times \Sigma \rightarrow [0,1]$ , une matrice indiquant les probabilités d'émission associées aux états ; on note  $P(y_t|q_t)$  la probabilité d'émettre à l'instant  $t$  le symbole  $y_t$  à partir de l'état  $q_t$ .
- $\pi: S \rightarrow [0, 1]$  est un vecteur de probabilités initiales de visite.

### *Exemple 1*

On dispose de 3 urnes contenant des boules blanches, noires et grises dans des proportions différentes. On note par  $P(B/U_i)$ ,  $P(N/U_i)$ ,  $P(G/U_i)$  les probabilités de tirer une boule blanche, noir ou grise de l'urne  $U_i$  ( $i=1, 2,3$ ).

Un opérateur **invisible** choisi, à des instants réguliers, une urne et tire une boule de cette urne et annonce la couleur. S'il choisi  $T$  urnes il annoncera (à un observateur) la séquence des  $T$  couleurs tirés notée  $O_T$ .

Ainsi, au temps  $t$ , l'opérateur choisi une urne  $U_t$  et tire une boule de couleur  $O_t$  de cette urne. On suppose que le tirage d'une boule au temps  $t$  est indépendant du tirage de la boule au temps  $t-1$ .

On fait l'hypothèse que l'opérateur choisi la séquence des urnes suivant un modèle d'une **chaîne de Markov** ayant pour paramètres  $\{\Pi, A\}$  et dont les états sont les urnes. Cette chaîne sera dite **cachée** (puisque l'opérateur est invisible).

**Ainsi :**

La séquence  $U_T=U_1U_2\dots U_T$  sera dite la **séquence des états cachée** (car invisible) et elle est générée par la chaîne de Markov cachée.

La séquence  $O_T=O_1O_2\dots O_T$  sera dite la **séquence des observations** qui sera annoncée à l'observateur.

### *Exemple 2*

L'exemple a pour objectif de définir à quelle saison nous sommes. Posons :

- Les 4 saisons (Printemps, Eté, Automne, Hiver), comme étant les états de notre modèle.

L'état "saison" n'est pas directement observable mais il émet des observations le temps de la journée. Il est défini par Nuage, Pluie, Soleil.

- Une chaîne d'observations le temps de la semaine. Dans les exemples cette chaîne est réduite à 3 jours pour limiter le nombre de calcul (Soleil, Soleil, Nuage) ou (S, S, N).

Le Modèle de Markov nous permettra de définir quelle est la chaîne de Markov qui a le plus de probabilité d'avoir généré la séquence observée, par exemple Eté, Eté, Printemps (E, E, P) Représentation graphique de ce modèle : Figure 3.3.

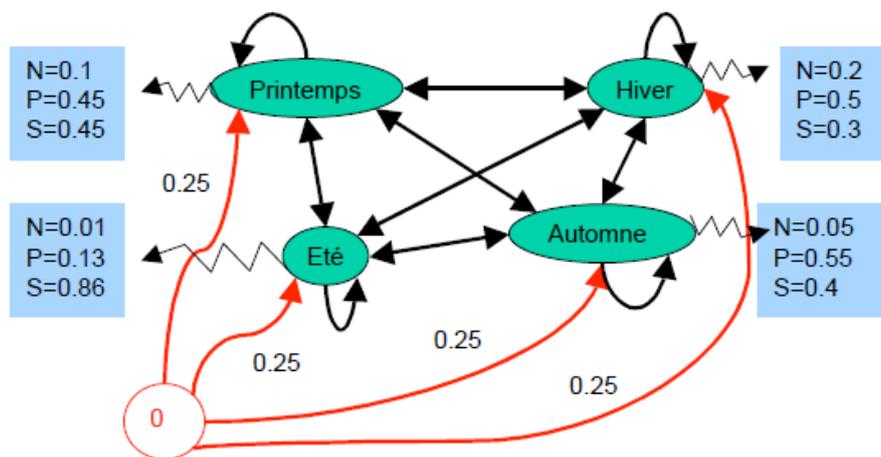


Figure 3.3: Modélisation HMM de l'exemple

## 2.2.2. Extensions des HMMs

### 2.2.2.1. Densités de probabilités

Selon le type de densité de probabilité d'observations, discrète ou continue, il est possible de construire deux types de modèles de HMM: soit un **HMM discret** soit un **HMM continu** [46].

- **HMM discret «Discrete Hidden Markov Models (DHMM) »**

**HMM discret «Discret Hidden Markov Models (DHMM) »** : est le modèle discret de Markov Caché et qui prend les chaînes des symboles discrets comme son entrée. Chaque symbole représente une classe des vecteurs caractéristiques. Il y a plusieurs méthodes pour obtenir ces symboles à partir d'un ensemble des vecteurs, autrement dit, il existe déjà

plusieurs méthodes afin de classifier cet ensemble en classes par exemple: K-moyennes (kmeans).

Dans les HMM discret les symboles d'observations en général sont continus puisqu'elles proviennent de phénomènes physiques continus. Ces observations continues sont quantifiées à l'aide d'un dictionnaire (symboles d'observations dénombrables).

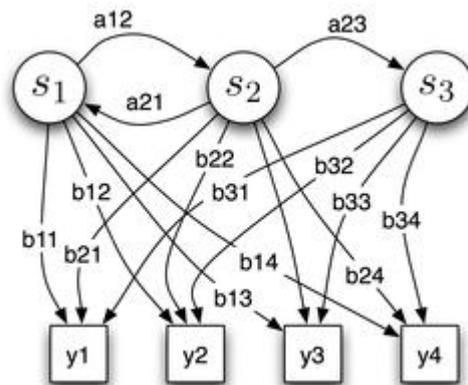


Figure 3.4 : Représentation d'un HMM discret à 3 états.

#### ➤ HMM continue « Continuous Hidden Markov Models (CHMM) »

Bien qu'il soit possible de quantifier les observations continues, il peut y avoir une sérieuse dégradation d'information associée à cette quantification. Il sera, alors, avantageux de choisir une fonction de densité de probabilités d'observations continues, conditionnée par les états du processus.

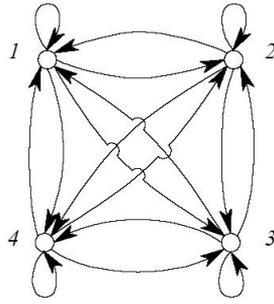
L'implémentation des HMMs continu est difficile et lente par contre l'implémentation des HMMs discret est Plus facile et plus rapide .et le nombre de paramètres à estimer dans CHMMs est élevé que DHMMs. [46].

### 2.2.3. Différents types de HMM

Différentes topologies de HMM peuvent être envisagées, les plus courants sont :

#### 2.2.3.1. Modèle ergodique

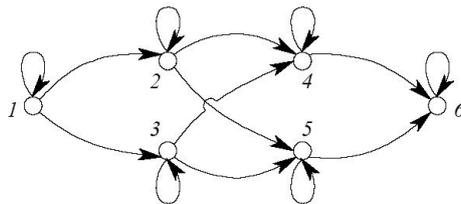
Celui ou chaque état peut être atteint à partir de n'importe quel autre en un nombre fini de transitions (Figure 3.5).



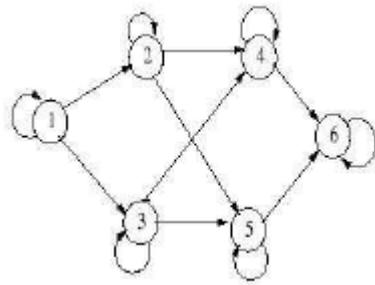
**Figure 3.5 :** *Modèle de chaîne de Markov ergodique.*

### 2.2.3.2. *Modèle gauche-droite*

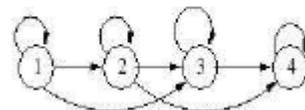
Une structure couramment utilisée est celle des modèles de Bakis, dits également modèles gauche-droite (figure 3.6), C'est un modèle où il y a des contraintes sur des transitions : seulement la transition d'un état ayant un indice bas vers un état ayant un indice haut est acceptée. Il existe deux sous-types de ce modèle : le modèle parallèle et le modèle séquentiel.



**Figure 3.6 :** *Modèle de Bakis.*



**3.6.1. Parallèle**



**3.6.2. Séquentiel**

Le modèle séquentiel est souvent utilisé pour reconnaître la parole et l'écriture manuscrite car il est convenable à la modélisation des signaux et il y a moins de paramètres à estimer dans la phase de l'apprentissage. [47]

#### 2.2.4. Les trois problèmes fondamentaux pour les HMMs

L'utilisation des HMMs nécessite la résolution des trois problèmes fondamentaux suivants :

- problème d'évaluation.
- problème de reconnaissance.
- Problème d'apprentissage.

On va détailler ces problèmes et présenter les algorithmes les plus utilisés pour les résoudre.

##### 2.2.4.1. Problème d'évaluation

Etant donné une suite d'observations  $O = \{O_1, O_2, \dots, O_m\}$  et un modèle  $\lambda$ , comment peut-on calculer efficacement la probabilité que la suite d'observations  $O$  soit produite par  $\lambda$ , c'est-à-dire  $P(O|\lambda)$  [48].

La solution à ce problème **d'évaluation** est l'algorithme Forward-Backward, qui fournit une solution exacte à ce problème faisant intervenir tous les chemins dans le modèle HMM.

##### ➤ Algorithme forward-backward

Dans cette approche, on considère que l'observation peut se faire en deux étapes [48]:

**Forward:** L'émission de la suite d'observations  $\{O_1, O_2, O_t\}$  et la réalisation de l'état  $S_t$  au temps  $t$ .

**Backward:** L'émission de la suite d'observations  $\{O_{t+1}, O_{t+2}, \dots, O_T\}$  en partant de l'état  $S_t$  au temps  $t$ .

##### 2.2.4.2. Problème de reconnaissance

Etant donné un HMM  $\lambda$  et une séquence observée  $O$ , comment connaître la séquence des états cachés du  $\lambda$  qui a le plus probablement généré  $O$  [48]?

La solution à ce problème **de reconnaissance** est l'algorithme de Viterbi.

##### ➤ Algorithme de Viterbi

C'est un algorithme récursif qui permet de trouver à partir d'une suite d'observations, une solution optimale au problème d'estimation de la suite d'états. Il fournissant une solution approximative faisant intervenir uniquement le meilleur chemin dans le modèle HMM (trouver la meilleure séquence d'états).

#### ***2.2.4.3. Problème d'apprentissage***

Etant donné un HMM  $\lambda = \{\Pi, A, B\}$  et une séquence observée  $O$ , Comment peut-on ajuster les paramètres du modèle  $\lambda$  pour maximiser la probabilité  $P(O|\lambda)$  [48] ?

La solution à ce problème **d'apprentissage** est l'algorithme de Baum-Welch :  $P(O|\lambda)$  est estimé en tenant compte de tous les chemins possibles (implémentation de l'algorithme Expectation-maximisation (EM)).

##### ***➤ Algorithme de Baum-Welch***

L'algorithme Baum-Welch est basé sur le théorème de Baum qui garantit l'atteinte d'un maximum local de la fonction de la probabilité par ré-estimation des paramètres  $A, B, \Pi$ . [49] Cette méthode de Maximum de la probabilité est la plus utilisée dans les applications.

#### ***2.2.5. L'utilisation des modèles de Markov cachés***

Les Modèles de Markov Cachés (HMM) sont utilisés avec succès depuis longtemps dans le domaine de la reconnaissance de la parole [48]. Ils ont ensuite été largement utilisés dans différents domaines tels que, notamment, l'analyse de séquences biologiques (notamment de gènes, [50]). De l'image [51]. Et de la vision par ordinateur [52] Ainsi, dans le domaine de la vision par ordinateur, les HMM ont pu être utilisés pour la reconnaissance de mouvements [53], [54]. Ils ont aussi été largement considérés pour la reconnaissance de contenu dans des vidéos de Sport [55]. Les HMM ont également été appliqués à l'indexation vidéo multimodale, plus précisément en combinant des informations tant visuelles que sonores [56].

### ***3. La quantification vectorielle***

Les vecteurs de données issus des mesures ont des composantes réelles. Pour les traiter sous la forme d'un modèle de Markov caché, il faut les quantifier. Le but fondamental de la quantification vectorielle est le partitionnement de l'espace des vecteurs de caractéristiques en

un certain nombre de groupes ou classes de façon à assurer une similarité intra-classe maximale et inter-classe minimale.

La quantification vectorielle (QV) [57] consiste à représenter tout vecteur  $x$  de dimension  $k$  par un autre vecteur  $y_i$  de même dimension, mais ce dernier appartenant à un ensemble fini  $\Delta$  de  $L$  vecteurs. Les vecteurs  $y_i$  sont appelés vecteurs représentants, vecteurs de reproduction ou code-vecteurs.  $\Delta$  est le dictionnaire ou le catalogue des formes. Nous représentons un vecteur  $x$  de l'espace  $\mathbb{R}^k$  par une matrice colonne  $x(n)$  :  $x=(x(1), x(2), \dots, x(k))^T$  où  $(.)^T$  indique l'opération de transposition d'une Matrice.

Un quantificateur vectoriel de dimension  $k$  et taille  $L$  peut être défini mathématiquement comme une application  $Q$  de  $\mathbb{R}^k$  vers  $\Delta$ .

$$Q : \mathbb{R}^k \rightarrow \Delta$$

$$x \rightarrow Q(x) = y_i$$

$$\text{avec } \Delta = \{y_i \in \mathbb{R}^k / i = 1, 2, \dots, L\}$$

Cette application  $Q$  détermine implicitement une partition de l'espace source  $\mathbb{R}^k$  en  $L$  régions  $C_i$ . Ces régions encore appelées classes sont déterminées par :  $C_i = \{x \in \mathbb{R}^k / Q(x) = y_i\}$

L'application  $Q$  associant à chaque vecteur d'entrée  $x$  de dimension  $k$  un vecteur  $y = Q(x)$  de même dimension (codeword) appartenant à un ensemble fini  $\Delta$  appelé dictionnaire (codebook) de taille finie  $L$ .

### Exemple

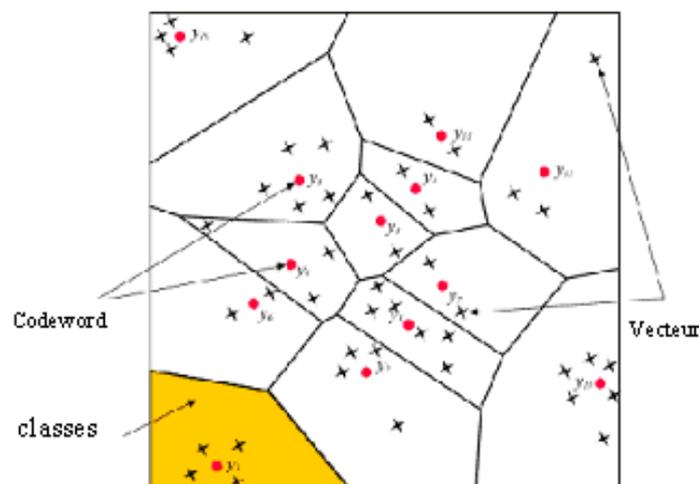


Figure 3.7: Quantification vectorielle,  $k=2$ ,  $L=13$

### 3.1. Algorithme de quantification

Il existe de nombreux algorithmes de quantification vectorielle, comme l'algorithme de K-means.

#### ➤ *k-means*

Cet algorithme est le plus connu parmi les algorithmes de quantification vectorielle.

L'algorithme des K-Means est une méthode de classification automatique (clustering), on l'appelle également K-Moyennes ou encore Nuées dynamiques (bien que cette dénomination constitue en réalité une généralisation).

Le principe de cette méthode est le suivant :

- Choisir le nombre de classes créer  $k$ .
- Choisir  $k$  centres initiaux  $C_1, \dots, C_k$ .
- Pour chaque vecteur, l'affecter au classe dont le centre est le plus proche
- Si aucun vecteur ne change de classe
- Alors STOP
- Sinon
- Calculer les nouveaux centres :

✓ Pour  $i = 1$  à  $k$

$C_i$  est la moyenne des éléments de la classe

✓ Aller en 3.

Sa simplicité fait la popularité de cette approche. Autre avantage notable, il n'est pas nécessaire de calculer au préalable les distances deux à deux entre tous les individus, opération très gourmande en temps et en espace mémoire dans certains algorithmes de clustering.

## Exemple

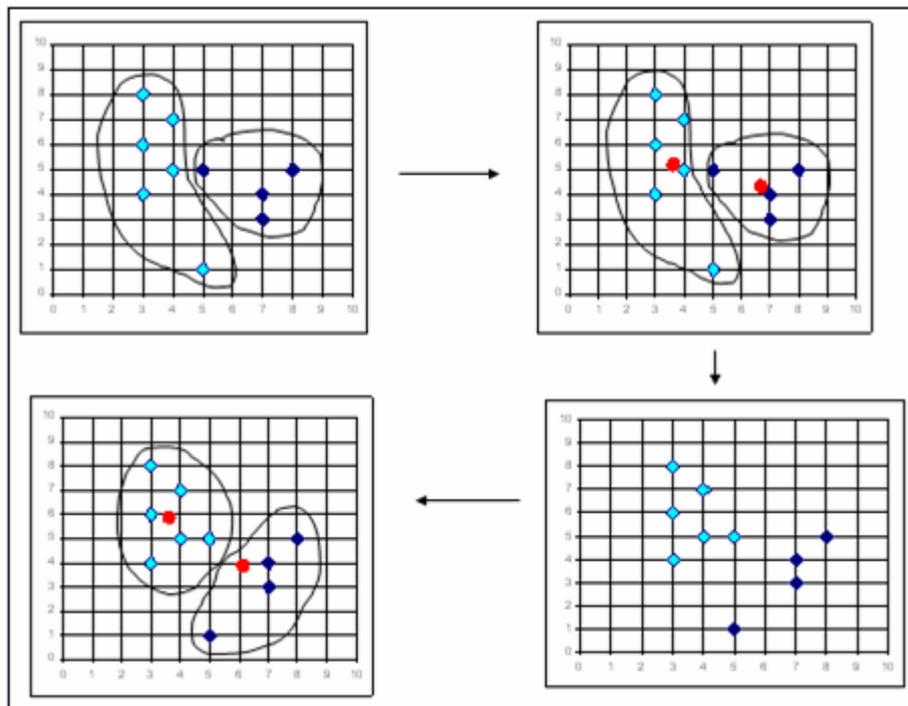


Figure 3.8 : K-means.

## 4. Conclusion

- Les HMMs sont des outils mathématiques puissants et très élégants qui peuvent être appliqués dans beaucoup de domaines.
- Leurs récentes extensions notamment au temps réel peuvent les rendre encore importants pour beaucoup de tâches d'apprentissage.
- Plusieurs problèmes d'ordres théoriques et pratiques restent à explorer, ce qui donne des challenges pour la recherche future.

## Chapitre 4 :

---

---

### Conception

---

---

#### 1. Introduction

Dans le domaine de la vision par ordinateur et plus particulièrement pour l'analyse et l'interprétation de mouvement humain, la reconnaissance de postures à partir de séquence vidéo est un problème d'actualité, Ce problème s'inscrit dans le champ de recherche plus général de l'interprétation de séquence vidéo.

Dans ce chapitre nous présentons la conception de notre système. Cette présentation englobe l'aspect architectural de l'application et les étapes relatives au domaine de la vision d'ordinateurs et plus spécifiquement à la reconnaissance de postures humaines.

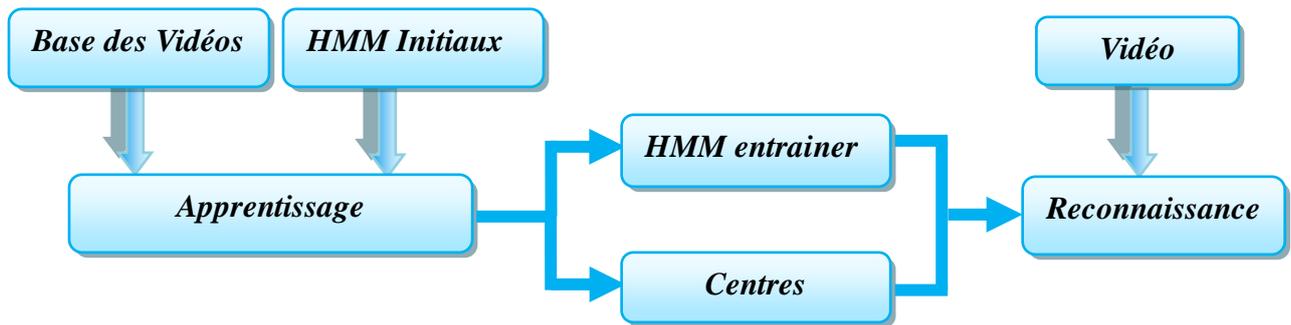
#### 2. Présentation générale

Notre projet se compose en deux parties : une partie d'apprentissage et une partie de reconnaissance.

Dans l'apprentissage on traite un ensemble des vidéos. Tous d'abord on commence par la lecture des vidéos et pour chaque vidéo nous appliquons plusieurs traitements: d'abord l'extraction des vecteurs de caractéristique par le calcul de flot optique, ensuite la Quantification vectorielle en utilisant l'algorithme de k-means pour classer l'ensemble des vecteurs obtenus. Ils représenteront les séquences d'observations pour l'apprentissage des HMMs. Cette classification nous donne un ensemble de classes, où les vecteurs de la même classe représentent la même observation. Les centroides de chaque classe sont sauvegardés. L'apprentissage est fait par l'algorithme de Baum-Welch.

Dans la partie reconnaissance, les vidéos en entrée subissent les mêmes traitements que celles utilisées dans l'apprentissage. On commence par la lecture d'une vidéo et l'extraction des caractéristiques, mais maintenant pour chaque vecteur on calcule la distance euclidienne par rapport aux centres enregistrés précédemment dans la phase d'apprentissage et on les classent dans la classe la plus proche, ce qui donne l'observation en entrée aux HMMs. Ensuite la reconnaissance est faite par l'algorithme de viterbi.

La figure 4.1 illustre l'implémentation générale du système.



*Figure 4.1: Vue générale de l'implémentation*

### **3. Partie d'apprentissage**

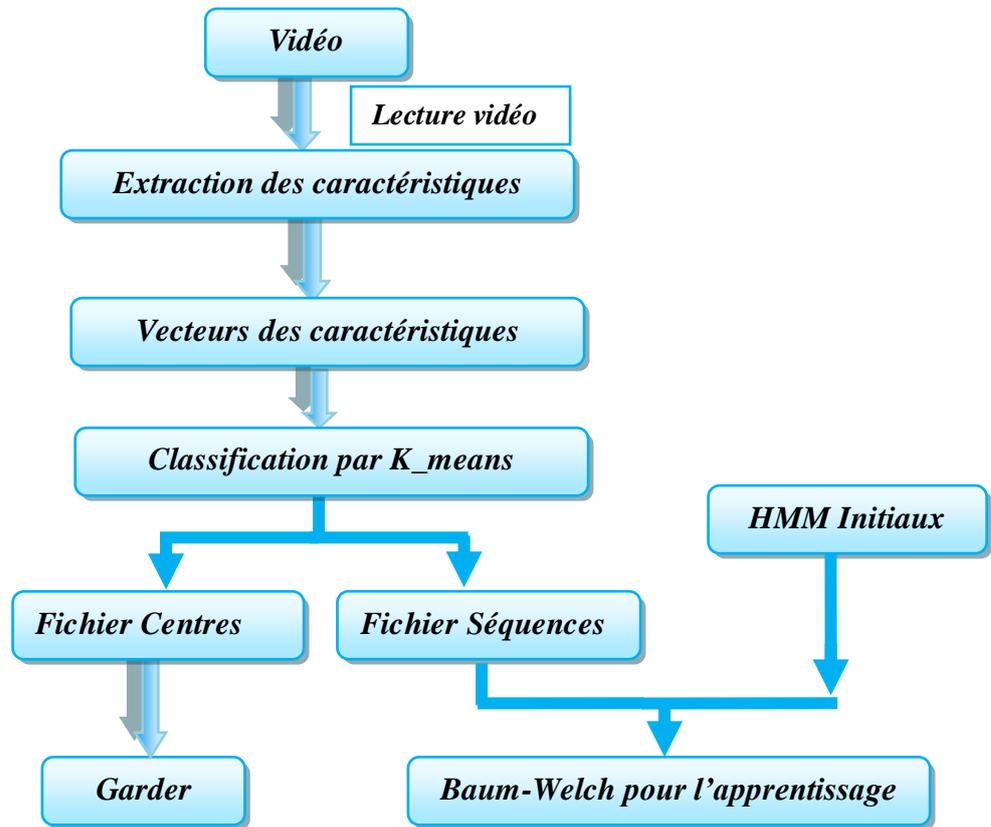
Dans cette partie on traite un ensemble des vidéos afin de train le système, cette phase est basé sur l'utilisation de l'algorithme de Baum-Welch. La sortie de cette étape est HMM entraîner.

Pour cela les traitements effectués se résument à :

- Une base des vidéos est préalablement construite. Cette base contient un ensemble des vidéos.
- Lecture des fichiers vidéos.
- Extraction des vecteurs de caractéristiques par le calcul de l'histogramme du flot optique.
- Quantification vectorielle des vecteurs de caractéristique par la méthode de K-moyennes (K-Means).
- Sauvegarde des séquences et des centroïdes.
- Soumission des séquences d'observations aux HMMs initiaux.
- L'apprentissage des HMM par l'algorithme de Baum-Welch.

En sortie on obtient des HMMs entraînés et les centres des classes d'observations pour la quantification vectorielle.

La figure 4.2 illustre schématiquement le processus de la partie apprentissage.



*Figure 4.2 : grandes étapes de la phase apprentissage des HMM.*

### *3.1. Construction de la base des vidéos*

*La* base de vidéos provient essentiellement de celle du *Carnegie Mellon University Motion database*. Accessible au : <http://mocap.cs.cmu.edu/>

*La* base contient une grande panoplie de vidéos touchantes un large spectre des mouvements humains. Nous nous sommes limités au mouvement de marche et saut. Le but étant de mettre en œuvre un système basé sur les HMMs.

### *3.2. Lecture de la vidéo*

*Dans* cette partie, on lit nos données qui sont une séquence vidéo. Les vidéos dans notre application ont certaines caractéristiques :

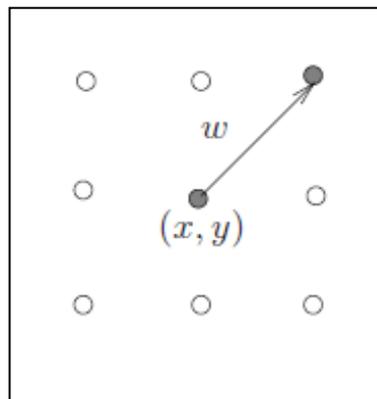
- Une Caméra Fixe.
- Un Fond Statique.
- Un seul objet pour traiter à la fois.

*Les* traitements suivants sont effectués sur tous les frames de la vidéo.

### *3.3. Extraction des vecteurs de caractéristiques*

Après la lecture de la vidéo, on passe à l'étape d'extraction des vecteurs de caractéristiques, la procédure choisie pour réaliser cette opération est l'histogramme de flot optique.

Le calcul de flot optique est appliqué entre deux images successives d'une séquence vidéo, l'utilisation du flux optique procure une information de mouvement pour chaque pixel de l'image. Ainsi, il mesure les vecteurs de déplacement à partir de l'intensité des pixels de deux images consécutives ou temporellement rapprochées. Dans un contexte de détection de mouvement, les pixels inactifs posséderont alors une vélocité nulle contrairement aux pixels appartenant à des objets dynamiques.



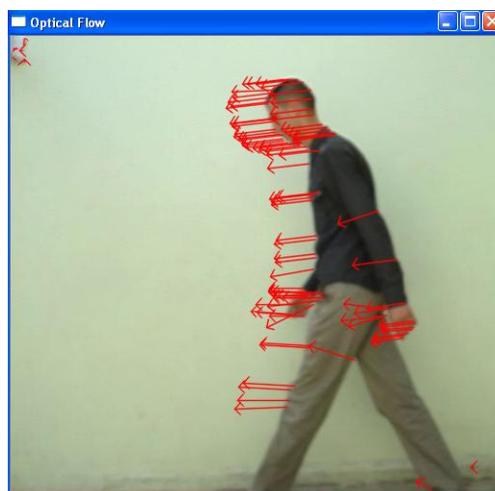
**Figure 4.3:**Transport de la luminosité

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t), \forall (x, y) \in \Omega.$$

$\Omega$ : domaine de l'image (région bornée de  $\mathbb{R}^2$ ).

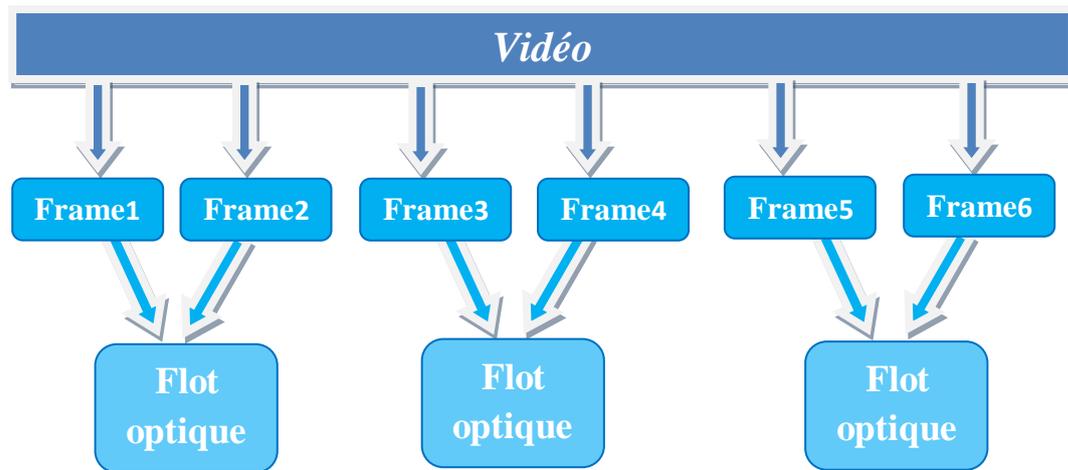
$w = (\delta x, \delta y)$  : déplacement du point  $(x, y)$  au temps  $t$ .

La Figure 4.4 illustre un exemple de calcul du flot optique d'une personne qui marche.



**Figure 4.4 :** le flot optique d'une personne qui marche

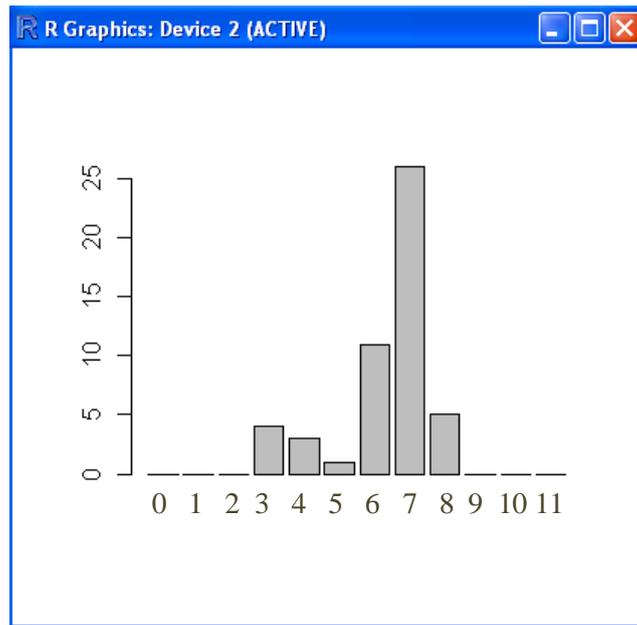
Un exemple de notre processus de calcul de flot optique est présenté dans la figure 4.5.



**Figure 4.5:** Un exemple de notre processus de calcul de flot optique

Un vecteur Flot Optique est caractérisé par un Angle et une Longueur.

Un histogramme est calculé sur la base des angles obtenus par le calcul du flot optique. La figure suivante illustre l'histogramme d'une frame du mouvement marche.



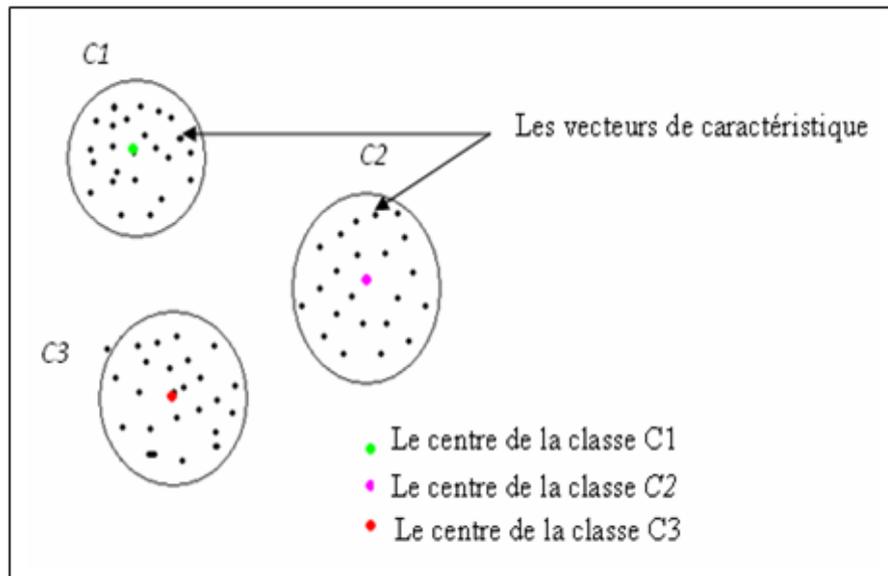
*Figure 4.6: l'histogramme d'une frame du mouvement marche*

### *3.4. Quantification vectorielle (Classification des vecteurs de caractéristique par K-Means)*

Après l'extraction des vecteurs de caractéristique par la procédure de l'histogramme de flot optique, nous avons appliqué une quantification vectorielle dont le but est de partitionner l'espace des vecteurs de caractéristiques en un certain nombre de groupes ou classes. L'objectif est de profiter des avantages des HMMs discrets.

La technique choisie pour faire la quantification est celle des K-Means, dans ce type d'algorithme la classe est représentée par son centre de gravité.

Un exemple sur la quantification par K-Means est illustré par la figure 4.7.



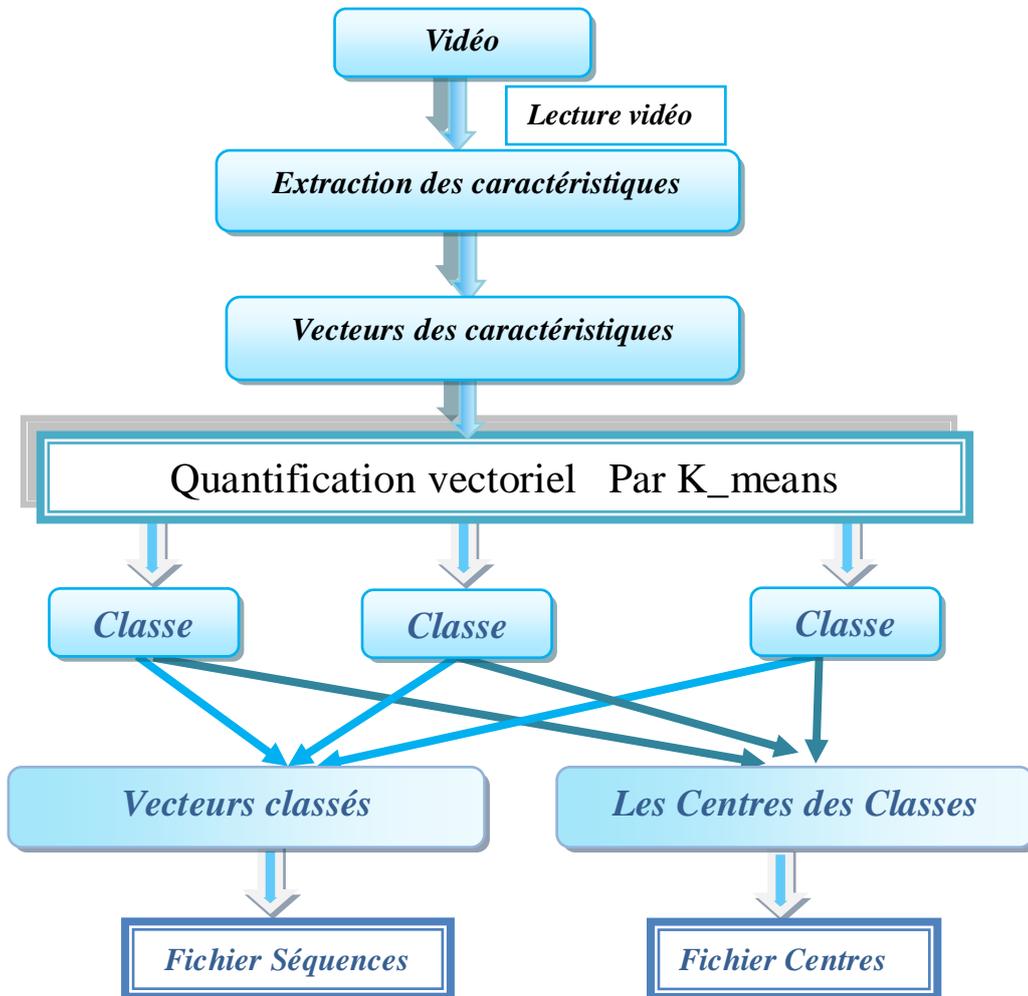
*Figure 4.7: Exemple sur la quantification par K-Means.*

### *3.5. Les résultats de la Classification (Quantification vectorielle)*

L'application de l'algorithme des K-Means sur tous les vecteurs de caractéristiques de toutes les vidéos de la base va générer la classification de ces vecteurs. Nous avons mis les résultats de cette étape dans deux fichiers:

- un fichier qui contient les séquences des symboles d'observation, Ce fichier va servir comme entrée des HMMs (l'algorithme de Baum-Welch).
- Et l'autre contient des centres. Ce fichier garde les centres de chaque classe pour les utiliser dans la deuxième partie.

La figure 4.8 illustre schématiquement le processus de la Classification.

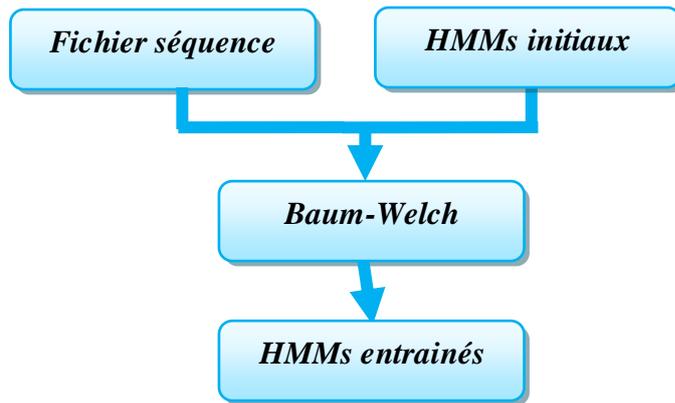


*Figure 4.8 : Le processus de la Classification*

### *3.6 L'apprentissage par les HMM (l'algorithme de Baum-Welch)*

Les Chaines de Markov Caché ont été largement appliquées pour résoudre un grand nombre de problèmes. Ils ont été adoptés dans le champ de recherche d'activité humain. Il peut être appliquée de manière efficace pour la reconnaissance des activités humaines. Les Chaines de Markov Caché est un modèle statistique dans lequel les états sont reliés à un vecteur d'observation. Dans notre cas, un état correspond à un mouvement.

Après la génération du fichier texte qui contient les séquences, ce fichier sera utilisé par l'algorithme de **Baum-Welch** avec les HMM initiaux qui contiennent les probabilités initiales d'émission entre les états et les observations et les probabilités initiales de transition entre les états. Les sorties de cet algorithme sont les HMM entraînés.

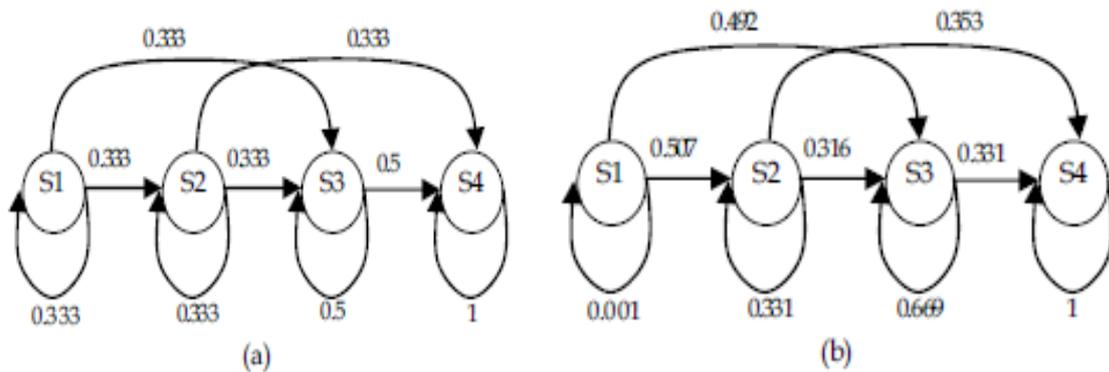


**Figure 4.9:** l'algorithme de Baum-Welch

### 3.7. Les résultats d'HMM après l'entraînement (Baum-Welch)

Les sorties de l'algorithme de Baum-Welch sont les HMMs entraîné qui contiennent les probabilités d'émission et de transition finales. C'est ces HMMs qui vont être utilisés pour la reconnaissance.

La figure 4.10 illustre un exemple d' HMM avant et après l'entraînement.



**Figure 4.10:** les probabilités de transition d'un HMM (a) avant et (b) après entraînement.

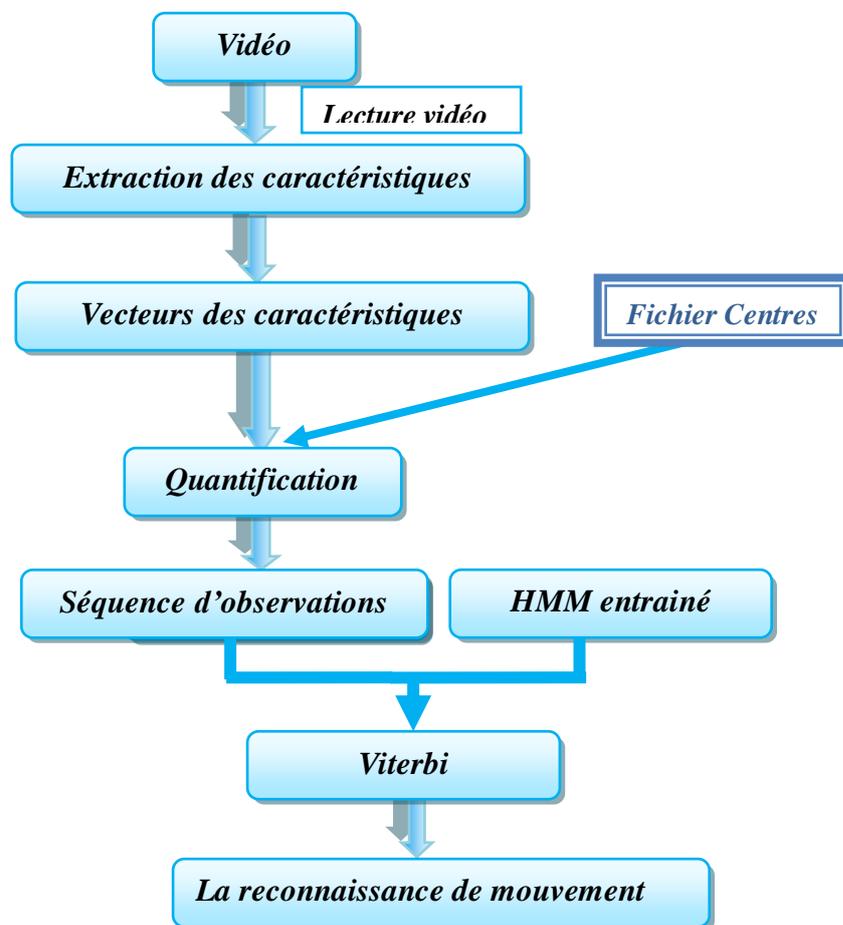
## 4. Partie reconnaissance

Dans cette partie on s'intéresse à reconnaître les gestes fait par la personne dans la vidéo, elle est fondées sur les résultats de la première partie. Pour chaque vidéo d'entrée les traitements suivants sont effectués.

- Lecture de fichier video.
- Capture les frames de la vidéo.
- L'extraction des caractéristiques par le flot optique.

- Pour chaque vecteur de caractéristique faire :
  - ✓ Une quantification vectoriel par le calcul de la distance euclidienne entre le vecteur de caractéristique et les centres des classes enregistrées précédemment (le résultat de k\_means dans la phase d'apprentissage).
  - ✓ Soumettre les séquences d'observations aux HMMs.
  - ✓ La reconnaissance par l'algorithme de viterbi.

La figure 4.11 illustre le processus de la partie de reconnaissance.



*Figure 4.11 : le processus de la reconnaissance de mouvement*

#### *4.1. La quantification vectoriel*

L'objectif de la quantification est de classer un élément en entrée par rapport à un corpus préétablis de classes. Dans notre cas il s'agit de classer le vecteur extrait de chaque frame de la

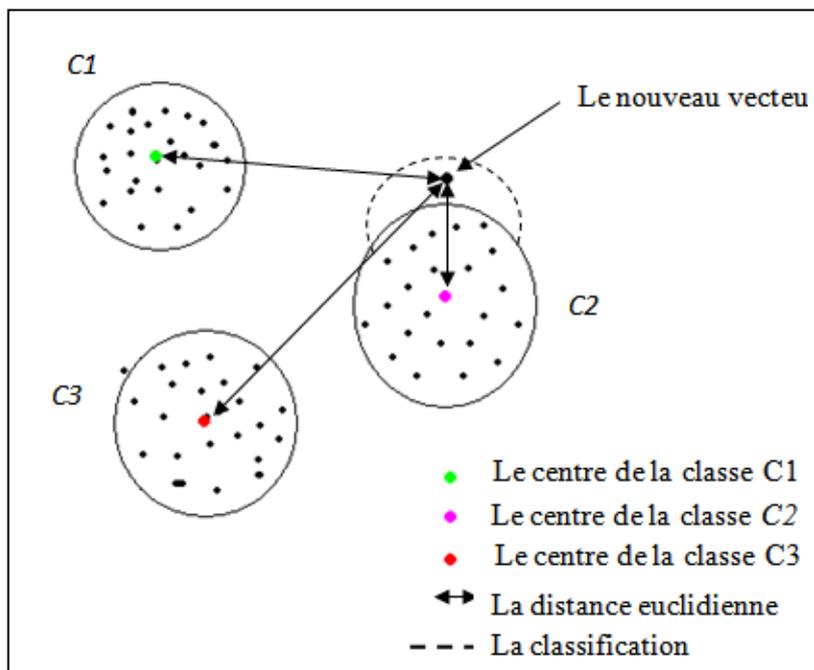
vidéo par rapport à l'ensemble des centres des classes qui ont été calculés précédemment par la méthode de classification k-means dans la partie apprentissage.

En d'autre terme, pour chaque vecteur de frame, il faut trouver la classe à laquelle il appartient.

La méthode choisie est celle du plus proche voisin sur une base de distance euclidienne. Les étapes successives sont :

- calcule la distance euclidienne entre le vecteur d'entrée et chaque centroïde.
- Si la distance est minimale entre le vecteur et le centre de la classe alors classé ce vecteur dans cette classe et mettre le résultat dans un fichier texte nommé séquences qui contient les symboles d'observation.

Un exemple qui exprime le principe de la distance euclidienne dans la figure 4.12.



*Figure 4.12 : le principe de la distance euclidienne*

Soit  $X = (x_1, \dots, x_n)$  et  $Y = (y_1, \dots, y_n)$  deux vecteurs, la distance euclidienne entre X et Y est:

$$D(x_i, y_i) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

## 4.2. La reconnaissance par les HMM (l'algorithme de Viterbi)

Après la quantification, on arrive à l'étape de la reconnaissance, Cette dernière se fait par l'algorithme de Viterbi.

Après l'entraînement par l'algorithme de Baum-Welch, les HMM entraînés seront utilisés par l'algorithme de VITERBI avec le fichier de séquences (résultat de la quantification) pour reconnaître les mouvements d'un être humain dans une vidéo, les résultats sont mis dans un fichier texte comme le montre la Figure 4.13.

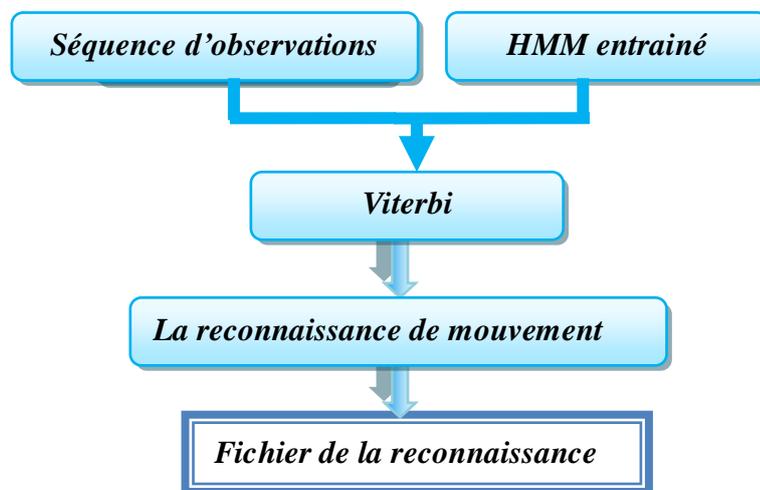


Figure 4.13 : Les entrées et les sorties de l'algorithme de viterbi

## 5. Conclusion

Dans ce chapitre on a présenté la conception détaillée du travail qu'on a réalisé pour atteindre notre objectif.

On a débuté par la phase d'apprentissage où nous avons créé une base des vidéos. Les opérations sont appliqués sur ces vidéo sont: la lecture, l'extraction des vecteurs de caractéristique ; ce dernier est fait par le calcul de l'histogramme de flot optique. Ensuite la quantification des vecteur par k\_means, après la classification nous avons met leurs résultats dans un fichier de séquence ce dernier sera l'entrée de l'algorithme de Baum-Welch avec les HMM initiaux, les résultats de Baum-Welch seront utilisés dans la phase de reconnaissance par l'algorithme de Viterbi.

Les détails de l'implémentation et les résultats obtenus seront présentés dans le prochain chapitre.

## Chapitre 5 :

---

# L'implémentation

---

### 1. Introduction

On arrive à la dernière étape dans notre mémoire, l'implémentation. Dans ce chapitre on présentera les techniques et outils utilisés pour l'implémentation du système. Ensuite nous allons exposer les différentes étapes de la réalisation de notre projet qui à pour objectif la réalisation d'un système de reconnaissance de mouvement et postures d'une personne dans une vidéo puis nous illustrons quelques résultats obtenus.

### 2. Les logiciels utilisés

#### 2.1. Code Block

Le Code :: Blocks est une open source environnement de développement intégré (IDE) multiplateforme (Linux, Mac et Windows) entièrement codé en C++. cela aide les développeurs dans la compilation et le débogage C et C ++.

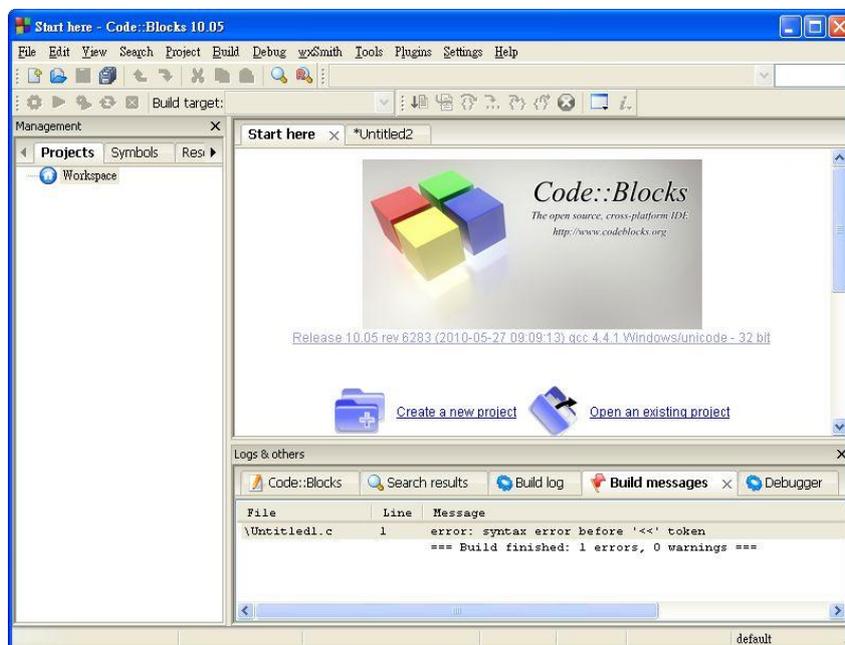


Figure 5.1 : l'interface de code block

## 2.2. Open CV

*Il* s'agit d'une bibliothèque créée à la base par la société Intel, qui propose un ensemble d'objets et de fonctions en liés à la vision par ordinateur.

*Elle* a pour mérite d'être gratuite, libre et multiplateformes, ce qui va nous permettre de l'utiliser aussi bien sous Windows, sous GNU/Linux, ou sous Mac OS.

*Elle* se divise en 5 sous-bibliothèques aux intérêts distincts :

- **Cv** contient les fonctions de base pour gérer et traiter nos images et vidéos ; Traitement image et outils de vision (Opérations morphologiques, Filtrage, Transformations géométriques, Histogramme, Détection de caractéristiques, Détection de mouvement et Suivi d'objets, Analyse structurelle, Détection d'objets (Classifieur de Haar) et Calibration de caméra et reconstruction 3D).
- **Cv aux** est un ensemble d'algorithmes dépréciés ou destinés à être intégrés à cv dans l'avenir, mais encore "à l'essai" ; Algorithmes expérimentaux (fonctions obsolètes) (Modèle de Markov 1D 2D, reconnaissance gestuelle (stéréo vision), description de texture, Tracking visage, Tracking 3D, Trouver squelette, Vidéo surveillance ... peu documenté).
- **Cxcore** contient les structures de base, ainsi que les fonctions d'algèbre linéaire indispensables ; Définition des Structures, Manipulations/opérations matricielle. Dessin (création de formes), Manipulation XML, Algorithme de Clustering (Classification), Macros (gestion d'erreur, mémoire).
- **Highgui** permet de créer des fenêtres dans lesquelles nous pourrions afficher nos images, même dans une application "console", et de gérer les captures vidéo.
- **ML** regroupe de nombreuses fonctionnalités autour de l'apprentissage ("Machine Learning"). Fonctions d'apprentissage, Fonctions de classification (outils de clustering)

*L'*un des objectifs Open CV est de fournir une infrastructure simple à utiliser la vision par ordinateur qui aide les gens à bâtir des applications de vision assez sophistiqués rapidement.

*La* bibliothèque Open CV contient plus de 500 fonctions qui couvrent de nombreux domaines de la vision, y compris l'inspection des produits en usine, l'imagerie médicale, la sécurité, l'interface utilisateur, étalonnage de la caméra, la vision stéréoscopique, et la robotique.

*Parce* que la vision par ordinateur et l'apprentissage machine vont souvent main dans la main, Open CV contient également aussi une Bibliothèque d'Apprentissage de Machine universelle

(Machine Learning Library) ou (MLL).

Cette sous bibliothèque est axé sur la reconnaissance des modèles statistiques. La MLL est très utile pour les tâches de vision qui sont au cœur de la mission Open CV, mais il est assez général pour être utilisé pour n'importe quelle machine d'apprentissage par problème.

Cette bibliothèque s'est imposée comme un standard dans le domaine de la recherche parce qu'elle propose un nombre important d'outils issus de l'état de l'art en vision des ordinateurs tels que :

- détection de droites, de segment et de cercles par Transformée de Hough
- détection de visages par AdaBoost
- cascade de classifieurs boostés
- détection de mouvement, historique du mouvement
- poursuite d'objets par meanshift ou Camshift
- détection de points d'intérêts
- estimation de flux optique (tracker de Kanade-Lucas)
- triangulation de Delaunay
- diagramme de Voronoi
- enveloppe convexe
- ajustement d'une ellipse à un ensemble de points par la méthode des moindres carrés

Certains algorithmes classiques dans le domaine de l'apprentissage artificiel sont aussi disponibles :

- k-means
- AdaBoost
- réseau de neurones
- machine à vecteurs de support
- Estimateur (statistique)

La figure 5.2 illustre un vue générale de l'ensemble de Open CV.

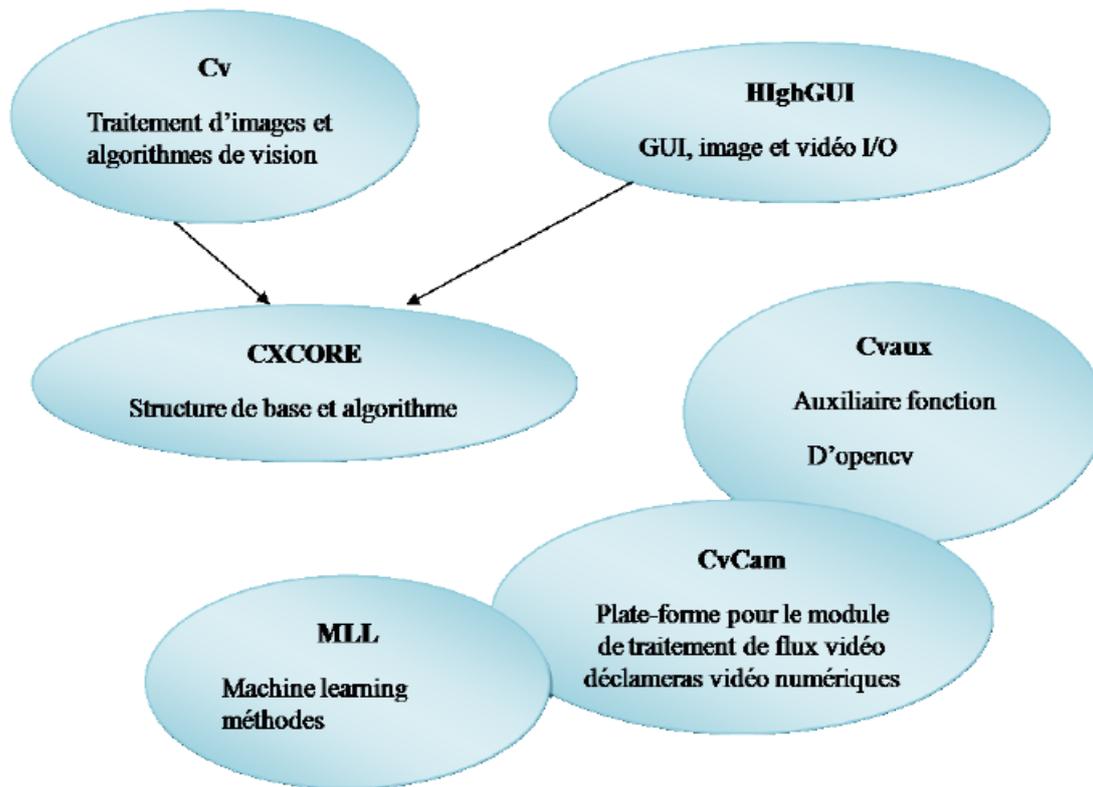


Figure 5.2 : Vue d'ensemble d'opencv

### 3. Présentation de l'application

Nous allons maintenant détaillée l'application réalisée.

#### 3.1. Base des vidéos

Nous avons utilisé les versions animées de la base CMU :

<http://mocap.cs.cmu.edu/search.php?subjectnumber=7&motion=%>

Chaque vidéo contient un mouvement (marche et saut). 20 vidéos ont été collectées. Parmi ces 20 vidéos, 08 vidéos ont été exploitées pour élaborer la base d'apprentissage et le reste ont été utilisées pour testé la reconnaissance. L'extension de ces vidéos est '.avi'.

La Figure 5.3 et 5.4 montre un échantillon des quelque frames de ces vidéos.

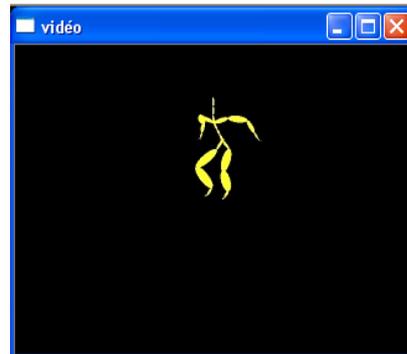
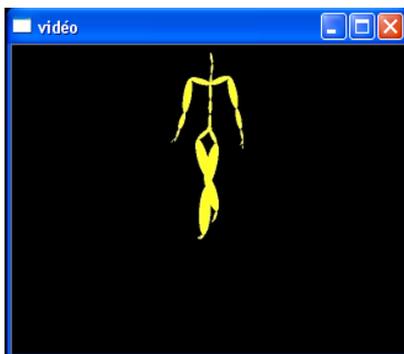


Figure 5.3 : frame d'une vidéo de la marche

Figure 5.4 : frame d'une vidéo de saut

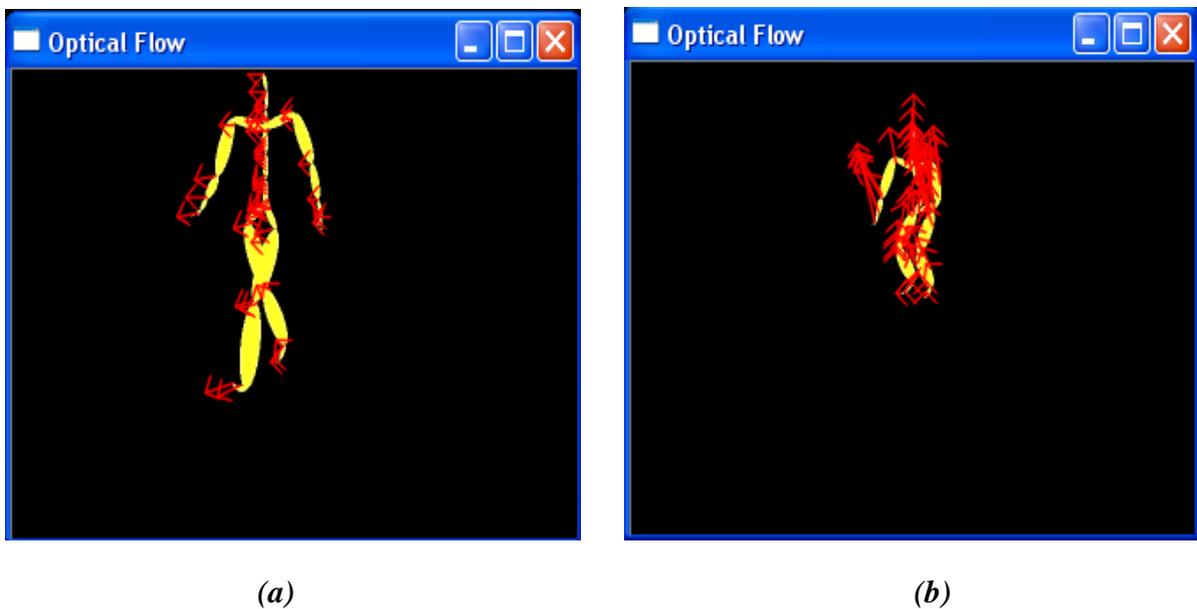
Nous avons filmé un ensemble des vidéos réelles afin de tester notre système.

### 3.2 Extraction des caractéristiques par le flot optique

Pour le calcul de flot optique nous avons utilisé le suivi de 100 points. Chaque point a un angle et une longueur, nous nous sommes limités à l'utilisation de l'information Angle qui semble apporter des indications sur les poses et mouvements du sujet.

L'espace de valeurs des Angles va de  $-\pi$  à  $\pi$ . Il est divisé en 12 intervalles dont l'histogramme (sur les 100 points) qui correspondront aux 06 symboles d'observation des HMMs.

La figure 5.5 illustre les vecteurs du flot optique de frames de la marche (a) et du saut (b).



**Figure 5.5 :** (a) Frame 18 d'une personne marche (b) Frame 12 d'une personne saute

Après nous avons calculé l'histogramme sur la base des angles obtenus.

Les vecteurs de caractéristique de ces deux frames sont:

<i>Les vecteurs de caractéristique</i>	<i>Histogramme d'angle</i>
<i>Frame d'une personne marche</i>	[0; 1; 1; 1; 3; 1; 16; 22; 2; 3; 0; 0]
<i>Frame d'une personne qui saute</i>	[0; 0; 13; 0; 19; 0; 14; 2; 0; 2; 0; 0]

**Tableau 5.1 :** Les vecteurs de caractéristique des quelque Frames de la vidéo

### 3.3. Quantification vectorielle :

Dans l'apprentissage, l'analyse de toutes les vidéos de la base génère un corpus de vecteurs. Ce corpus est soumis à l'algorithme K-Means pour classer chaque vecteur. Chaque classe de vecteurs sera représentées par son centre. Un vecteur (l'histogramme) appartenant à une classe est une émission du symbole de cette classe.

Les résultats de K-Means sont mis dans deux fichiers texte:

- le premier contient les séquences des symboles d'observation. par exemple, si le vecteur est classé dans la classe numéro 1 nous diront qu'on a observé le symbole C1, s'il est classé dans la classe2 c'est le symbole C2... etc. donc les données de ce fichiers sont ressemblent à :

C2 C2 C1 C1 C1 C3 C0 C1 C3 C0

C3 C1 C4 C1 C1 C3 C1 C4 C1 C3

C2 C1 C4 C1 C3 C3 C0 C3 C2 C2

- Le deuxième fichier contient des centres. Par exemple si on a 3 classes :

<i>N<sup>o</sup> de classe</i>	<i>Les centres</i>
<i>01</i>	[3.4705882, 44.411766, 2.1176472]
<i>02</i>	[0.23529412, 17.058823, 32.529411]
<i>03</i>	[41.799999, 7.5999999, 0.40000001]

**Tableau 5.2 : les centres des classes**

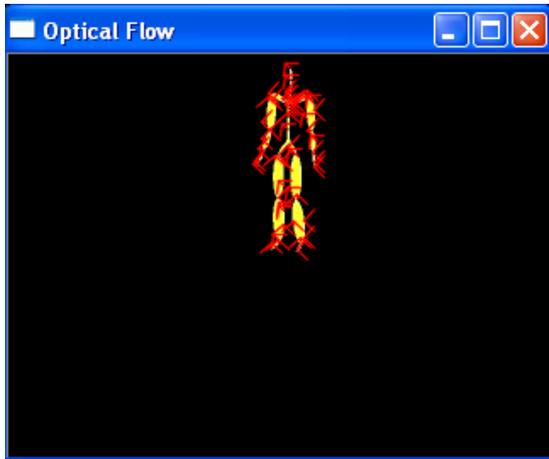
### 3.4. Les chaine de Markov caché

Pour l'implémentation des HMMs nous avons utilisé un package qui se trouve dans : <http://webdocs.cs.ualberta.ca/~lindek/hmm.htm> et nous avons appliqué dans notre cas qui est la reconnaissance des postures et mouvement.

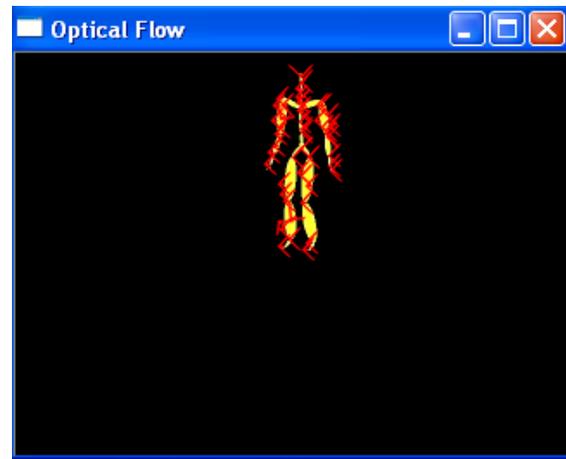
Chaque mouvement est composé de plusieurs poses, nous avons fait la reconnaissance sur deux mouvements humains (saute et marche) où on considère que chaque mouvement est composé d'états qui sont les poses. Le choix de ces poses est fait par une observation sur les histogrammes des frames, nous considérons que les frames qui possède les mêmes histogrammes appartiennent à la même classe de poses (état).

Selon les histogrammes nous concluons que le mouvement de saute est composé de 4 poses. L'étude des histogrammes permet de construire les HMM adéquats et de les initialiser.

- **La Pose1**: se compose d'un ensemble des frames d'un homme debout (à l'arrêt) comme:

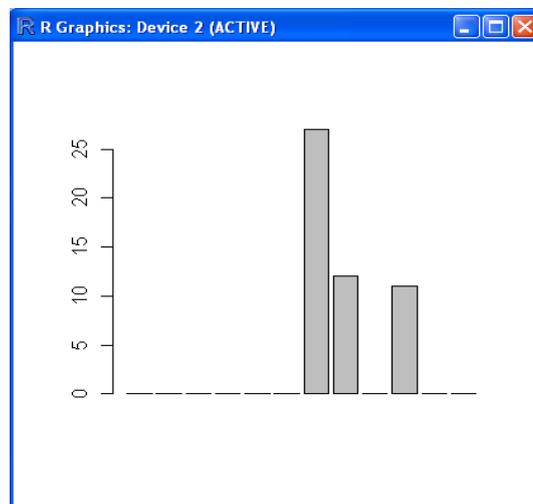


*Frame0*



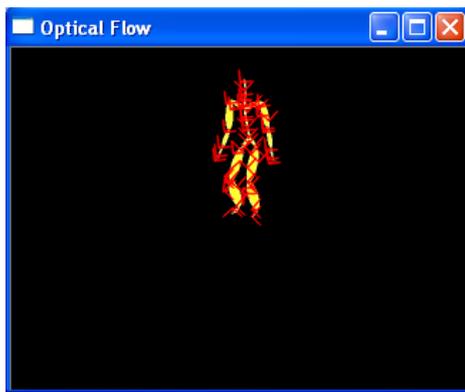
*Frame8*

Les frames de cette pose entre la frame0 et frame8 ont presque tous des histogrammes comme suit:

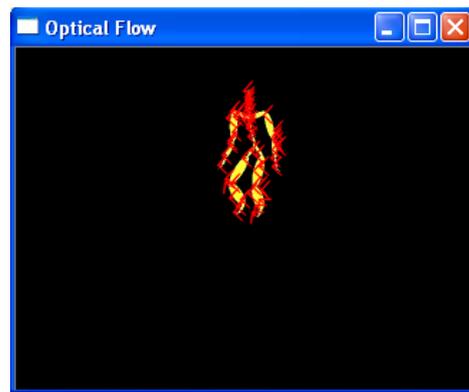


*Figure 5.6 : L'historgramme des frames entre 0 et 8*

- **Pose2:** se compose d'un ensemble des frames d'un homme qui amorce le saut:

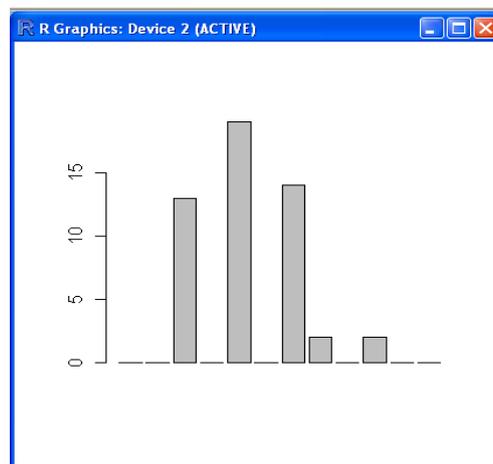


*Frame9*



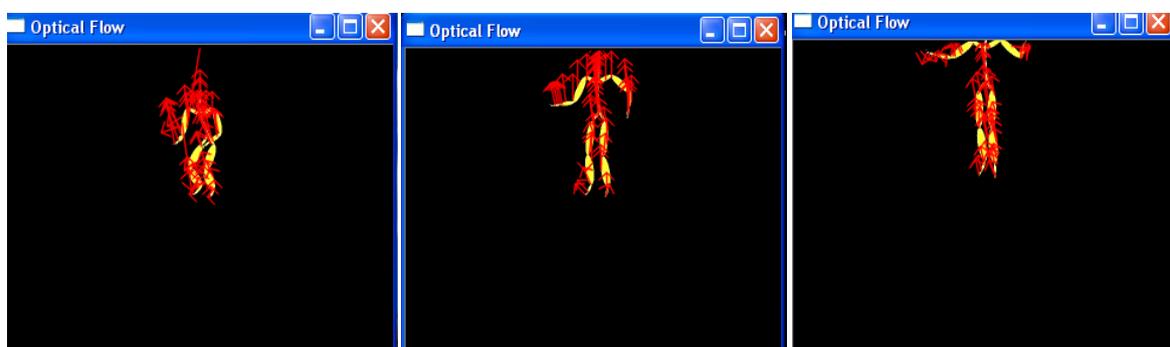
*Frame14*

Les frames de cette pose entre la frame9 et frame14 ont presque les mêmes histogrammes comme suit:



*Figure 5.7: L'histogramme des frames entre 9 et 14*

- **Pose3:** se compose d'un ensemble des frames d'un homme qui part vers le haut du saut:



*Frame15*

*Frame18*

*Frame23*

Les frames de cette pose entre la frame15 et frame23 ont presque les mêmes formes d'histogrammes comme suit:

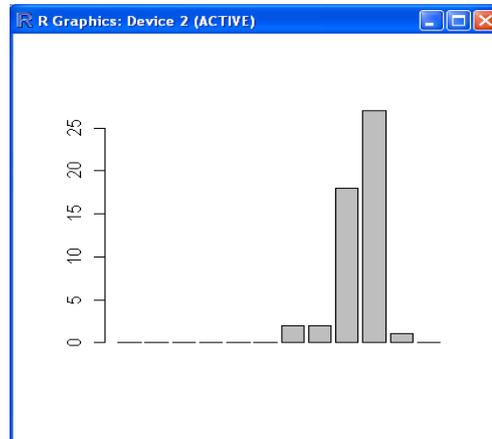
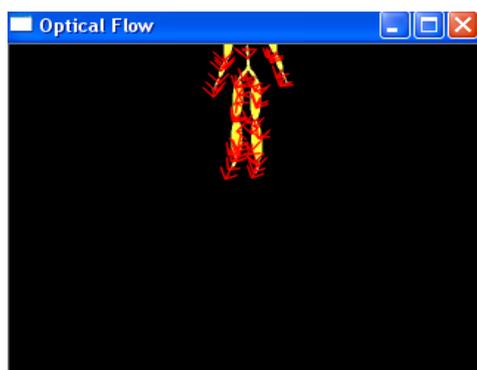
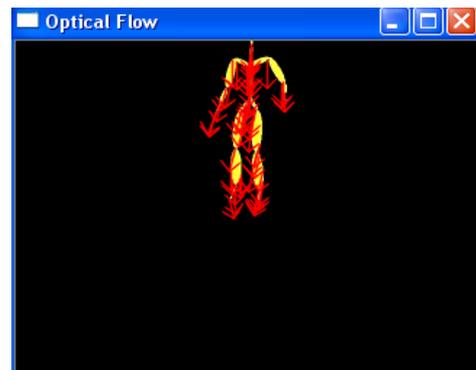


Figure 5.8 : L'histogramme des frames entre 15 et 23

➤ **Pose4** : se compose d'un ensemble des frames d'un homme en phase descende:



Frame27



Frame31

Les frames de cette pose entre la frame27 et frame 31 ont presque les mêmes histogrammes comme suit:

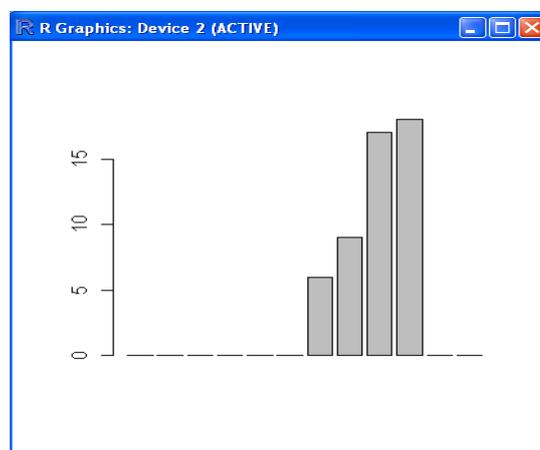


Figure 5.9 : L'histogramme des frames entre 27 et 31

### 3.4.1. L'apprentissage

Pour l'apprentissage nous avons utilisé l'algorithme de **Baum-Welch** qui prend en entrée les séquences d'observation et les HMMs initiaux qui sont alimentés par la séquence d'observation issue des vidéos de la base.

Nous avons expérimenté notre méthode sur deux exemples. Le premier illustre la décomposition du mouvement en poses (deux HMMs : figure 5.10 et 5.11), le deuxième tente de reconnaître le mouvement directement à partir des observations (figure 5.12).

#### 1<sup>er</sup> CAS : Observation Angles → Poses → Mouvements

##### ➤ Les HMMs avant et après l'entraînement

✓ Les probabilités de transition entre les états:

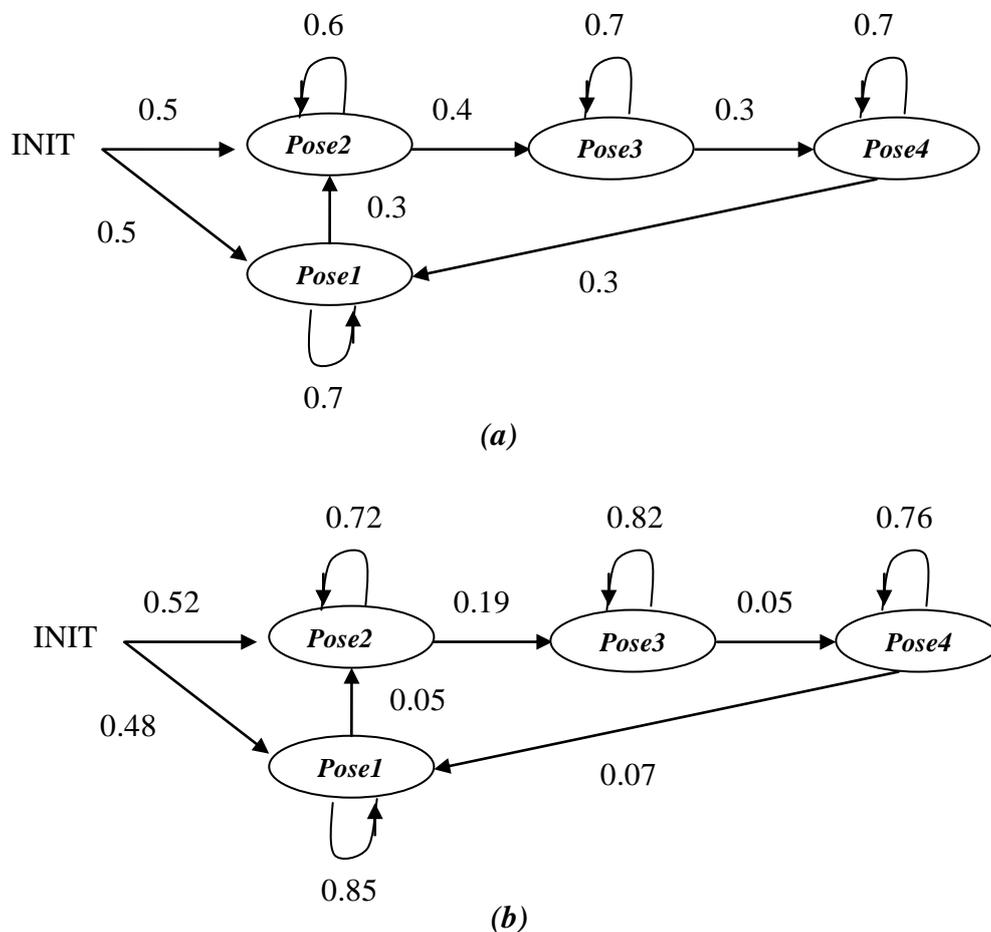
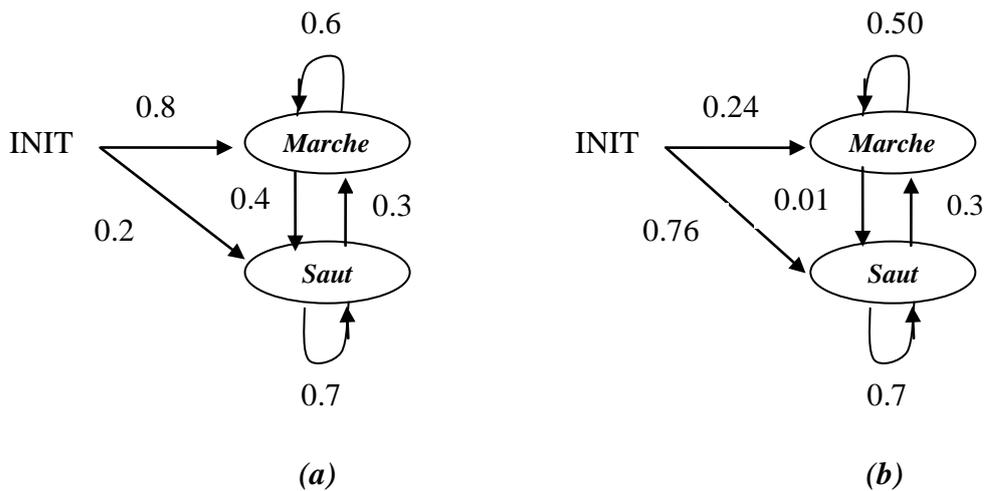


Figure 5.10 : Les probabilités de transition d'un HMM (poses)(a) avant et (b) après l'entraînement



**Figure 5.11** : Les probabilités de transition d'un HMM (mouvement globale) (a) avant et (b) après l'entraînement

✓ **Les probabilités initiales d'émission entre les états et les observations :**

- Les émissions entre les poses et les observations (Ci) :

POSE1	C2	0.45
POSE1	C1	0.45
POSE1	C5	0.1
POSE2	C3	1
POSE3	C4	1
POSE4	C0	1

- Les émissions entre les mouvements et les observations (poses) :

MARCHE	POSE1	1
SAUT	POSE1	0.25
SAUT	POSE2	0.25
SAUT	POSE3	0.25
SAUT	POSE4	0.25

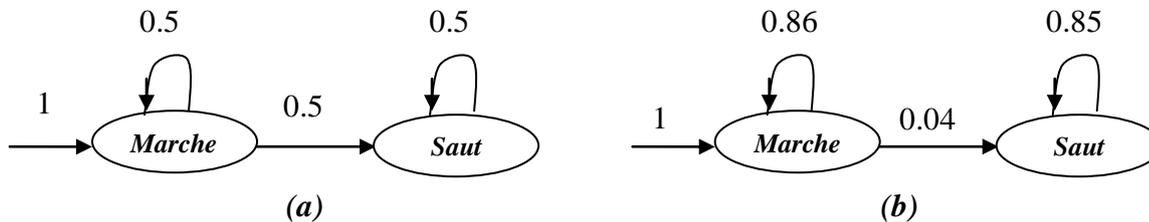
Notez que l'HMM de la figure 5.10 prend les symboles Ci comme observations et produits des états POSEi. L'HMM de la figure 5.11 prend comme observations les états produits par le premier HMM.

**2ém CAS : Observation Angles → Mouvements**

Ce HMM reconnait le mouvement directement à partir des observations sur les histogrammes d'angles du flot optique (Ci).

➤ *Les HMMs avant et après l'entraînement*

✓ *Les probabilités de transition entre les états:*



**Figure 5.12:** Les probabilités de transition d'un HMM (a) avant et (b) après l'entraînement

**3.4.2. La reconnaissance**

Pour reconnaître les mouvements d'un humain nous avons utilisé l'algorithme de VITERBI qui prend en entrée:

- Les séquences d'observations (résultat de la quantification vectorielle qui est faite par le calcul de la distance euclidienne des centres des classes données par l'application du K-MEANS).
- L'HMM entraîné.

**4. Les résultats expérimentés**

Dans cette partie, on va aborder les résultats de la reconnaissance des poses et mouvements (marche et saut), Nous avons fait des tests sur plusieurs vidéos. Nous ne présentons ici que quelques résultats des HMM sur les deux exemples. Le premier illustre la décomposition du mouvement en poses (HMM1), et le deuxième tente de reconnaître le mouvement directement à partir des observations(HMM2) .Les vidéos de la marche sont numérotées de 1 à 3 et les vidéos de saut sont numérotées de 4 à 6.Les vidéos 1 et 4 concernent des vidéos de la base d'apprentissage, 3 et 6 concernent des vidéos de personnes réelles.

➤ Un échantillon des résultats sur les vidéos de la marche sont comme suit :

```

GN "D:\MASTER2\PROJE... _ _ _ X
P(path)=0.970762
path:
c2 POSE1
c1 POSE1
c1 POSE1
c2 POSE1
P(path)=0.974635
path:
c2 POSE1
c1 POSE1
c2 POSE1
c2 POSE1
P(path)=0.970397
path:
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.471797
path:
c2 POSE1
c2 POSE1
c5 POSE1
c5 POSE1
P(path)=0.976759
path:
c1 POSE1
c1 POSE1
c2 POSE1
c2 POSE1
P(path)=0.98021
path:
c1 POSE1
c2 POSE1
c2 POSE1
c1 POSE1
P(path)=0.946228
path:
c5 POSE1
c1 POSE1
c2 POSE1
c2 POSE1
P(path)=0.875629
path:
c1 POSE1
c5 POSE1
c1 POSE1
c1 POSE1

```

Vidéo1

```

GN "D:\MASTER2\PROJET... _ _ _ X
P(path)=0.974165
path:
c2 POSE1
c1 POSE1
c2 POSE1
c1 POSE1
P(path)=0.97655
path:
c2 POSE1
c2 POSE1
c2 POSE1
c1 POSE1
P(path)=0.423252
path:
c2 POSE1
c1 POSE1
c5 POSE1
c5 POSE1
P(path)=0.976402
path:
c1 POSE1
c2 POSE1
c2 POSE1
P(path)=0.80101
path:
c2 POSE1
c2 POSE1
c1 POSE1
c5 POSE1
c2 POSE1
P(path)=0.976402
path:
c1 POSE1
c2 POSE1
c2 POSE1
c2 POSE1
P(path)=0.829661
path:
c1 POSE1
c2 POSE1
c1 POSE1
c5 POSE1
c5 POSE1
P(path)=0.40867
path:
c5 POSE1
c5 POSE1
c1 POSE1
c1 POSE1

```

Vidéo 2

```

GN "G:\PROJET FINAL\vit... _ _ _ X
P(path)=0.970397
path:
c1 POSE1
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.957593
path:
c2 POSE1
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.970397
path:
c1 POSE1
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.970397
path:
c1 POSE1
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.970397
path:
c1 POSE1
c1 POSE1
c1 POSE1
c1 POSE1
P(path)=0.818738
path:
c2 POSE1
c1 POSE1
c5 POSE1
c1 POSE1

```

Vidéo 3

Figure5.13 : Les résultats de l'HMM1 de pose sur les vidéos de la marche

```

GN "D:\MASTER2\PROJE... _ _ _ X
P(path)=0.701115
path:
POSE1 MARCHE
P(path)=0.701115
path:
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE

```

Vidéo1

```

GN "D:\MASTER2\PROJET... _ _ _ X
P(path)=0.701115
path:
POSE1 MARCHE
P(path)=0.701115
path:
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE

```

Vidéo 2

```

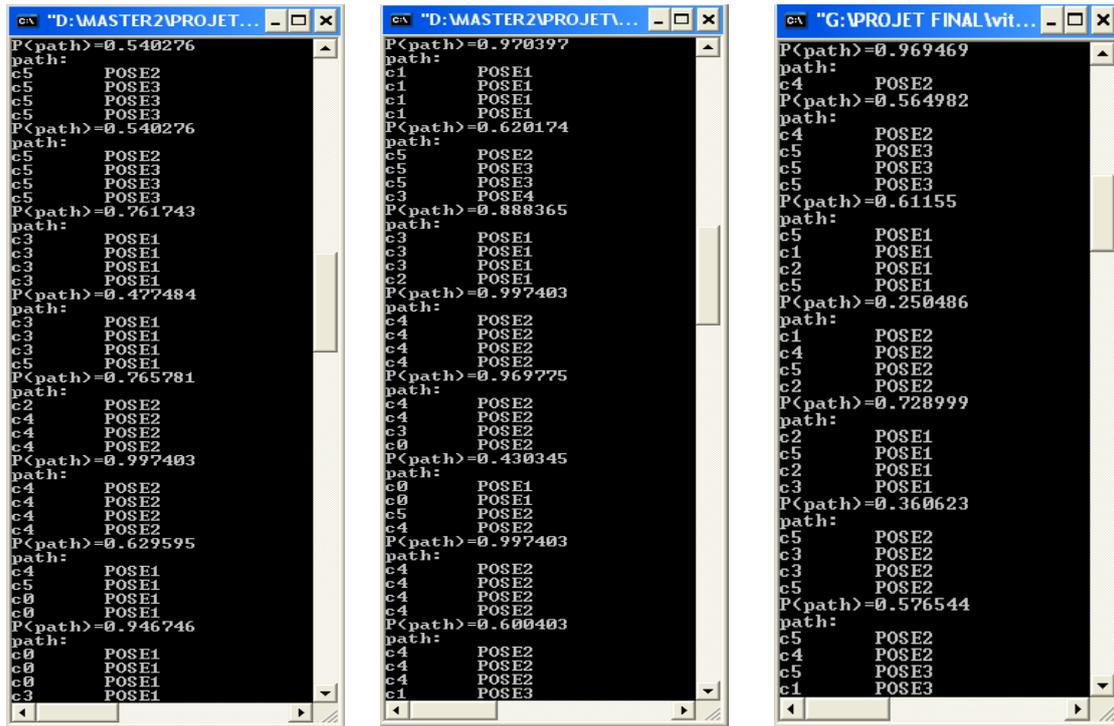
GN "G:\PROJET FINAL\vit... _ _ _ X
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
P(path)=0.817403
path:
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE
POSE1 MARCHE

```

Vidéo 3

Figure5.14 : Les résultats de l'HMM1 de mouvement sur les vidéos de la marche

➤ *Un échantillon des résultats sur les vidéos de saut sont comme suit :*

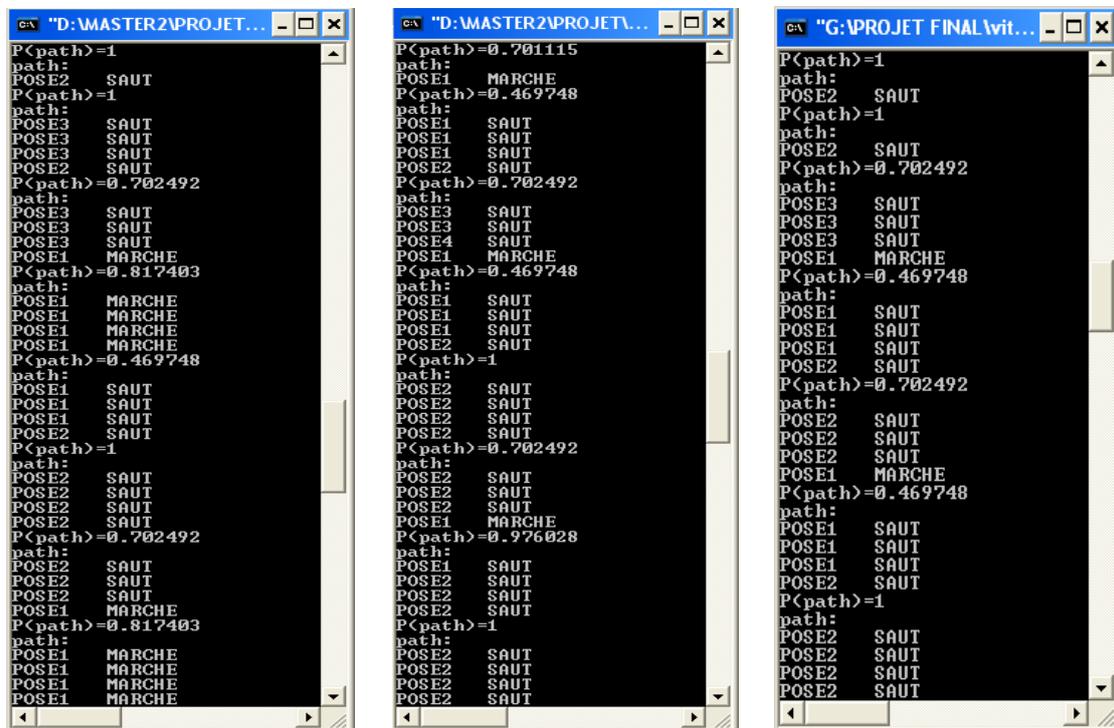


*Vidéo 4*

*Vidéo 5*

*Vidéo 6*

*Figure 5.15 : Les résultats de l'HMM1 de pose sur les vidéos de saut*



*Vidéo 4*

*Vidéo 5*

*Vidéo 6*

*Figure 5.16 : Les résultats de l'HMM1 de mouvement sur les vidéos de saut*

## 2ém CAS : Observation Angles → Mouvements

➤ *Un échantillon des résultats sur les vidéos de la marche sont comme suit :*

```

    "F:\PROJET2\witerbi ...
    P(path)=0.984206
    path:
    c9 MARCHÉ
    c7 MARCHÉ
    c7 MARCHÉ
    c7 MARCHÉ
    c7 MARCHÉ
    P(path)=0.978043
    path:
    c8 MARCHÉ
    c7 MARCHÉ
    c8 MARCHÉ
    c7 MARCHÉ
    c8 MARCHÉ
    P(path)=0.976119
    path:
    c9 MARCHÉ
    c7 MARCHÉ
    c2 MARCHÉ
    c2 MARCHÉ
    c9 MARCHÉ
    P(path)=0.975123
    path:
    c8 MARCHÉ
    c8 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    P(path)=0.975859
    path:
    c9 MARCHÉ
    c9 MARCHÉ
    c7 MARCHÉ
    c8 MARCHÉ
    c9 MARCHÉ
    P(path)=0.979792
    path:
    c9 MARCHÉ
    c8 MARCHÉ
    c2 MARCHÉ
    c2 MARCHÉ
    c8 MARCHÉ

    "F:\PROJET2\witerbi ...
    P(path)=0.982587
    path:
    c7 MARCHÉ
    c7 MARCHÉ
    c7 MARCHÉ
    c8 MARCHÉ
    c7 MARCHÉ
    P(path)=0.977951
    path:
    c9 MARCHÉ
    c9 MARCHÉ
    c8 MARCHÉ
    c7 MARCHÉ
    c8 MARCHÉ
    P(path)=0.975657
    path:
    c2 MARCHÉ
    c8 MARCHÉ
    c7 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    P(path)=0.987352
    path:
    c9 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    c8 MARCHÉ
    c2 MARCHÉ
    P(path)=0.982325
    path:
    c8 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    c7 MARCHÉ
    P(path)=0.987891
    path:
    c9 MARCHÉ
    c7 MARCHÉ
    c2 MARCHÉ
    c8 MARCHÉ
    c2 MARCHÉ

    "F:\PROJET2\witerbi ...
    P(path)=0.484796
    path:
    c5 MARCHÉ
    c2 MARCHÉ
    c1 MARCHÉ
    c0 MARCHÉ
    c2 MARCHÉ
    P(path)=0.988718
    path:
    c6 MARCHÉ
    c10 MARCHÉ
    c2 MARCHÉ
    c7 MARCHÉ
    c2 MARCHÉ
    P(path)=0.696294
    path:
    c2 MARCHÉ
    c2 MARCHÉ
    c2 MARCHÉ
    c2 MARCHÉ
    c1 MARCHÉ
    P(path)=0.519404
    path:
    c2 MARCHÉ
    c2 MARCHÉ
    c6 MARCHÉ
    c6 MARCHÉ
    c2 MARCHÉ
    P(path)=0.969978
    path:
    c8 MARCHÉ
    c0 MARCHÉ
    c2 MARCHÉ
    c8 MARCHÉ
    c10 MARCHÉ
    P(path)=0.979437
    path:
    c2 MARCHÉ
    c2 MARCHÉ
    c8 MARCHÉ
    c2 MARCHÉ
    c8 MARCHÉ
  
```

Vidéo 1

Vidéo 2

Vidéo 3

Figure5.17 : Les résultats de l'HMM2 de mouvement sur les vidéos de la marche

➤ *Un échantillon des résultats sur les vidéos de saut sont comme suit :*

```

    "F:\PROJET2\witerbi ...
    P(path)=0.979838
    path:
    c2 MARCHÉ
    c2 MARCHÉ
    c2 MARCHÉ
    c10 MARCHÉ
    c10 MARCHÉ
    P(path)=0.999172
    path:
    c5 MARCHÉ
    c5 SAUT
    c10 SAUT
    c2 SAUT
    c6 SAUT
    P(path)=0.999274
    path:
    c3 MARCHÉ
    c3 MARCHÉ
    c3 MARCHÉ
    c3 MARCHÉ
    c3 MARCHÉ
    P(path)=0.86772
    path:
    c6 MARCHÉ
    c2 MARCHÉ
    c5 SAUT
    c4 SAUT
    c11 SAUT
    P(path)=0.710682
    path:
    c11 MARCHÉ
    c4 SAUT
    c8 SAUT
    c6 SAUT
    c6 SAUT
    P(path)=0.885943
    path:
    c3 MARCHÉ
    c6 SAUT
    c6 SAUT
    c6 SAUT
    c6 SAUT

    "F:\PROJET2\witerbi ...
    P(path)=0.979332
    path:
    c2 MARCHÉ
    c2 MARCHÉ
    c10 MARCHÉ
    c10 MARCHÉ
    c10 MARCHÉ
    P(path)=0.641099
    path:
    c10 MARCHÉ
    c9 MARCHÉ
    c1 SAUT
    c1 SAUT
    c1 SAUT
    P(path)=0.847721
    path:
    c1 MARCHÉ
    c1 SAUT
    c6 SAUT
    c5 SAUT
    c11 SAUT
    P(path)=0.858202
    path:
    c11 MARCHÉ
    c11 SAUT
    c0 SAUT
    c1 SAUT
    c1 SAUT
    P(path)=0.831103
    path:
    c1 MARCHÉ
    c1 SAUT
    c1 SAUT
    c1 SAUT
    c1 SAUT
    P(path)=0.709167
    path:
    c6 MARCHÉ
    c7 MARCHÉ
    c7 MARCHÉ
    c0 SAUT
    c0 SAUT

    "F:\PROJET2\witerbi ...
    P(path)=0.959056
    path:
    c8 MARCHÉ
    c2 MARCHÉ
    c0 MARCHÉ
    c7 MARCHÉ
    c2 MARCHÉ
    P(path)=0.981451
    path:
    c8 MARCHÉ
    c5 SAUT
    c2 SAUT
    c9 SAUT
    c7 SAUT
    P(path)=0.999989
    path:
    c2 MARCHÉ
    c5 SAUT
    c5 SAUT
    c2 SAUT
    c5 SAUT
    P(path)=0.987352
    path:
    c8 MARCHÉ
    c2 MARCHÉ
    c9 MARCHÉ
    c9 MARCHÉ
    c2 MARCHÉ
    P(path)=0.984415
    path:
    c5 MARCHÉ
    c5 SAUT
    c2 SAUT
    c10 SAUT
    c8 SAUT
    P(path)=0.787565
    path:
    c2 MARCHÉ
    c0 SAUT
    c6 SAUT
    c2 SAUT
    c1 SAUT
  
```

Vidéo 4

Vidéo 5

Vidéo 6

Figure5.18 : Les résultats de l'HMM2 de mouvement sur les vidéos de saut

*Finalem*ent on peut conclure que, pour la mini base élaborée, les résultats de la reconnaissance sont assez satisfaisants. C'est-à-dire, les deux mouvements ont été bien reconnus (correctement). Cependant nous ne pouvons généraliser et plus de testes sont nécessaires.

## ***5. Conclusion***

*La* reconnaissance de posture et mouvement demande beaucoup de recherche, et reste toujours un domaine de recherche impressionnant et difficile à réaliser à 100%.

*Dans* notre application on a utilisé les logiciels et frame-work pour pouvoir atteindre un bon résultat, tels que l'éditeur de c++ code block ; les bibliothèques de vision par ordinateur OpenCV d'open source.

---

---

## *Conclusion générale*

---

---

*Dans ce mémoire, nous avons présenté un système de reconnaissance automatique des mouvements en utilisant les modèles de Markov cachés probabilistes, nous nous sommes intéressés à la reconnaissance des mouvements et postures d'un humain dans des scènes contraintes : mono caméra fixe, un seul sujet. Le système comporte deux grandes phases. La première permet de concevoir la base d'apprentissage par un processus qui adopte plusieurs étapes. La deuxième cherche à reconnaître les mouvements et les postures.*

*L'élaboration de la base d'apprentissage consiste à extraire les vecteurs de caractéristiques par le calcul de l'histogramme de flot optique, ensuite une classification se fait par la méthode du k-Means est appliqué sur les vecteurs obtenus. Et enfin l'apprentissage se fait par l'un des algorithmes des modèles de Markov cachés qui est l'algorithme de BAUM-WELCH.*

*Le processus de la reconnaissance, conçue pour reconnaître les mouvements d'un humain, intègre les résultats de la première partie. Cela est fait par le passage sur les mêmes premières étapes de l'apprentissage qui sont l'extraction des vecteurs de caractéristiques, ensuite une quantification vectorielle est faite par le calcul de la distance euclidien. Et enfin, un algorithme de VITERBI a été mise au point pour reconnaître des séquences d'états optimaux (le plus probable) dans un contexte d'états cachés et observations.*

*Pour l'apprentissage nous avons optés pour des vidéos de la base du CMU Graphics Lab Motion Capture Database. Nous avons choisis les séquences animées.*

*Le système réalisé a été testé sur un ensemble de vidéos de quelques mouvements. Selon les résultats obtenus plusieurs remarques peuvent être dégagées :*

- *L'objectif principal est de mettre en place (conception et implémentation) une méthode pour la reconnaissance de mouvements et postures.*
- *L'information donnée par les histogrammes d'angles issus de l'analyse du flot optique apparaît comme satisfaisante pour discerner entre certains mouvements.*
- *les résultats sont assez satisfaisants.*

*Il* faut néanmoins remarquer que l'apprentissage s'est fait avec un petit nombre de vidéos et donc les résultats pour l'instant ne permettent pas de trancher quand à l'efficacité de la méthode.

*Ce* travail nous a permis d'aborder d'une manière pratique plusieurs concept : HMM, Kmeans, Flot Optique, OpenCv, etc. le plus important reste pour nous l'intégration de la démarche scientifique en recherche.

*Enfin*, ce travail ouvre plusieurs perspectives résumées en ce qui suit :

- Mise en pratique de la technique pour une application réelle.
- Etendre l'approche aux scènes avec complexes : occlusion, plusieurs personnes.

---

---

## *Bibliographie*

---

---

- [1] <http://www.answers.com/topic/computer-vision>.2011
- [2] Bérard.F, « Vision par Ordinateur pour l'Interaction Homme-Machine Fortement Couplée ». Thèse de doctorat Informatique, Université Joseph Fourier-GRENOBLE I. mars 1992.
- [3] Dardar Selma, Kharoubi Madiha, «Reproduction de Mouvement Vidéo Par un Mime». Ingénierie des medias, 2010.
- [4] Yassine Boukir et Nadjib Benkhellat, « Segmentation d'images en niveaux de gris par des méthodes évolutives (AG, AGQ, ANTSEG) ». Mémoire d'ingénieur, INI 2005.
- [5] Mr. Slime Samir, « Environnement de segmentation d'image à base d'une approche biomimétique». Systèmes Informatiques. oued-smar alger : Mémoire d'ingénieur, 2007/2008, pages 134.
- [6] maizi djazira, khalla wafa, zeddouri samah, « approche segmentation hybride à base ». Intelligence artificielle, 2006.
- [7] S.Bres, J.Jolion et F.Lebourgeois, « Traitement et analyse des images numériques». Sermes Sciencepublication, 2003.
- [8] <http://www.optique-ingenieur.org/fr/cours> .2011.
- [9] N.Ayache and O. D.Faugeras, «HYPER.A new approach for the recognition and positioning of two\_dimensional objects». IEEE Trans.on Pattern Analysis and Machine Intelligence. PAMI\_ January 1986.
- [10] D.G. Lowe, « Organization and Visual Recognition ». Kluwer Academic-Publisher.1985.
- [11] Edelman, « Representation and Recognition in Vision». MIT Press, Cambridge, MA (1999).
- [12] Ilya Shlyakhter, Max Rozenoer, Julie Dorsey et Seth J. Teller « Re-constructing 3d tree models from instrumented photographs». IEEE Computer Graphics and Applications, 21(3):53-61, 2001.
- [13] L.Wang, W. Hu, T. Tan, «Recent Developments in Human Motion Analysis». National Laboratory of Pattern Recognition .Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080.1997.
- [14] R.T. Collins et al, «A system for video surveillance and monitoring» VSAM final report, CMU-RI-TR-00-12, Technical Report, Carnegie Mellon University, 2000.

- [15] Y. Cui, J.J. Weng, « Hand segmentation using learning-based prediction and verification for hand sign recognition». Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1997, pp. 88-93.
- [16] H.M. Lakany, G.M. Haycs, M. Hazlewood, S.J. Hillman, «Human walking: tracking and analysis». Proc. of IEE Colloquium on Motion Analysis and Tracking. 1999, pp. 5/1-5/14.
- [17] D. Meyer, J. Denzler and H. Niemann, «Model based extraction of articulated objects in image sequences for gait analysis». Proc. of IEEE Intl. Conf. on Image Processing. 1997, pp. 78-81.
- [18] J.J. Little, J.E. Boyd, «Recognizing people by their gait: the shape of motion». Videre: Journal of Computer Vision Research, The MIT Press, 1 (2), 1998.
- [19] D. Cunado, M.S. Nixon, J.N. Carter, «Automatic gait recognition via model-based evidence gathering». Proc. of Workshop on Automatic Identification Advanced Technologies. New Jersey, 1998, pp. 27-30.
- [20] W. Freeman et al, « Computer vision for computer games». Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. 1996, pp. 100-105.
- [21] K.P. Karmann, A. Brandt, « Moving object recognition using an adaptive background memory ».in V Cappellini, Time-varying Image Processing and Moving Object Recognition, 2.Elsevier, Amsterdam, The Netherlands, 1990.
- [22] M. Kilger, «A shadow handler in a video-based real-time traffic monitoring system». Proc. of IEEE Workshop on Applications of Computer Vision. 1992, pp. 1060-1066.
- [23] I. Haritaoglu, D. Harwood, L.S. Davis, W4, « real-time surveillance of people and their activities». IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.
- [24] R. Jain et H.-H. Nagel. «On the analysis of accumulative difference pictures from image sequences of real-world scenes». IEEE Trans. Pattern Analysis and Machine Intelligence, 1(2):206-214, 1979.
- [25] M. Kunert D. M. Gavrilu ET U. Lages, « A multi-sensor approach for the protection of vulnerable traffic participants ». The protector project. Dans Proc. of the IEEE Instrumentation and Measurement Technology Conference, volume 3, pages 2044–2048, 2001.
- [26] Alexandre Lemieux, « Systeme d'identification de personnes par vision numerique», These, Universite LAVAL, QUEBEC, Decembre 2003.

- [27] I.A. Karaulova, P.M. Hall, A.D. Marshall, «A hierarchical model of dynamics for tracking people with a single video camera». British Machine Vision Conference. 2000, pp. 352-361.
- [28] H. Sidenbladh, M. J. Black, D. J. Fleet, « Stochastic tracking of 3D human figures using 2D image motion». Proc. of European Conference on Computer Vision, 2000.
- [29] I.A. Kakadiaris, D. Metaxas, «Model-based estimation of 3-D human motion with occlusion based on active multi-viewpoint selection». Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, 1996, pp. 81-87.
- [30] T. Boult, «Frame-rate multi-body tracking for surveillance». DARPA Image Understanding Workshop, Monterey, Calif. San Francisco: Morgan Kaufmann, November 1998.
- [31] Christine LEIGNEL, J.M. JOLION, « Etude bibliographie suivi de personnes et trajectoire dans un réseau de caméras ». Rapport de recherche RR-LIRIS-2008-006.
- [32] Crowley J.L., Demazeau Y. « Principles and Techniques for Sensor Data Fusion, Signal Processing (EURASIP) ». Vol. 32, pages 5-27.
- [33] Chleq N., Thonnat M., « Realtime image sequence interpretation for vidéosurveillance». In IEEE, editor, International Conference on Image Processing, Lausanne, Switzerland, pages 801-804, 1996.
- [34] Baumberg A., Hogg D, « An adaptative eigenshape model». In British Machine Vision Conference BMVC, Birmingham, 1995.
- [35] Jain R., Martin W., Aggarwal J. « Segmentation throught the detection of changes due to motion In Computer Graphics and Image Processing», Vol. 2, pages 13-34.
- [36] Ricquebourg Y. «Segmentation et suivi d'objets mobiles par modèles structurels adaptatifs». Master's thesis, Institut national des sciences appliquées de Rennes, 1993.
- [37] T.B. Moeslund, E. Granum, « A survey of computer vision-based human motion capture». Computer Vision and Image Understanding 81 (3) (2001) 231–268.
- [38] M. Brand, Shadow puppetry, in: International Conference on Computer Vision, Corfu, Greece, Sep 1999.
- [39] G. Cheung, S. Baker, T. Kanade, « Shape-from-silhouette for articulated objects and its use for human body kinematics estimation and motion capture». in: Computer Vision and Pattern Recognition, Madison, Wisconsin, USA, June 16–22, 2003.
- [40] Y. Luo, T.-W. Wu, J.-N. Hwang, « Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks». Computer Vision and Image Understanding 92 (2003) 196–216.

- [41] C. Stauffer, W.E.L. Grimson, « Learning patterns of activity using real-time tracking». IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 747–757.
- [42] L. Wang, H. Ning, T. Tan, W. Hu, « Fusion of static and dynamic body biometrics for gait recognition». In: International Conference on Computer Vision, Nice, France, Oct 13–16, 2003.
- [43] J.W. Davis, S.R. Taylor, «Analysis and recognition of walking movements». In: International Conference on Pattern Recognition, Quebec, Canada, Aug 11–15, 2002.
- [44] I. Haritaoglu, D. Harwood, L.S. Davis, W4, «real-time surveillance of people and their activities». IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.
- [45] L. Wang, H. Ning, T. Tan, W. Hu, « Silhouette analysis-based gait recognition for human identification ». IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1505–1518.
- [46 ] J. Dequier, «Chaînes de Markov et applications». Examen probatoire en informatique du CNAM, centre régional RHÔNE-ALPES, centre d'enseignement de GRENOBLE, 2005.
- [47]Khalid Hallouli, « Reconnaissance de caractères par méthodes markoviennes et réseaux bayésiens ».Rapport de thèse, l'Ecole Nationale Supérieure des Télécommunications2004
- [48] L.R. Rabiner, «A Tutorial on Hidden Markov Models and Select Application in Speech Recognition». Proceeding of the IEEE, vol. 77, n° 2, pp.257-286, 1989.
- [49] A. Belaid, G. Saon, « Utilisation des processus markoviens en reconnaissance de l'écriture». Traitement du Signal, vol. 14, n°. 2, pp. 161 – 177, 1997.
- [50]R. Durbin, S. R. Eddy, A. Krogh, G. Mitchison, Biological Sequence Analysis : «Probabilistic Models of Proteins and Nucleic Acids». Cambridge University Press. 1998.
- [51]J. Li, A. Najmi, R. M. Gray, « Image classification by a two dimensional hidden Markov model».. IEEE Transactions on Signal Processing, 48(2) :517 :533, février 2000.
- [52]Starner T. Starner, « Visual recognition of American sign language using hidden Markov models». Master's Thesis, MIT, Février 1995.
- [53]J. Hoey et J. J. Little, «Representation and recognition of complex human motion». Proc. of the Int. Conf. on Computer Vision and Pattern Recognition, Hilton Head Island, juin 2000.

**[54]**M. H. Cheng, M. F. Ho et C. L. Huang, « Gait Analysis For Human Identification Through Manifold Learning and HMM». Pattern Recognition, 41(8) :2541-2553, août 2008.

**[55]** A. Kokaram, N. Rea, R. Dahyot, « Browsing sports video (Trends in sports-related indexing and retrieval work) ». IEEE Signal Processing Magazine, 23(2) :47-58, mars 2006.

**[56]**R. Leonardi et P. Miglioratti, « Semantic indexing of multimedia documents». IEEE Transactions on Multimedia, 9(2):44-51, avril 2002.

**[57]** Gersho et Gray, « vector quantization and signal compression Boston». kluwer academic publishers 1992.