

# Chapitre 6



## Modélisation d'humains

# Pourquoi la modélisation d'humains?

---

## □ Environnement immersif

- Objectif: Utiliser le traitement vidéo pour contrôler un environnement de façon non intrusive.

## ■ Exemple: KidsRoom

- Soustraction d'arrière-plan
- Composantes connectées pour former des régions correspondant aux enfants et à un lit.
- Modélisation des régions par leurs couleurs (couleur moyenne), dimensions et leur trajectoire.

# Pourquoi la modélisation d'humains?

## □ Environnement immersif (suite)

- Détection de l'intensité du mouvement (changement pixel par pixel).
- Détection de gestes (taille de régions, formes à l'aide de moments, patrons temporels et moments).



# Plan du chapitre

---

## ❑ Modélisation d'humains

### ■ Objectifs

### ■ Méthodes

#### ❑ Par la silhouette (contour)

#### ❑ Par régions

#### ❑ Par des bâtonnets

## ❑ Reconnaissance de démarche

# Modélisation d'humains

---

- ❑ Objectifs:
  - Suivi des humains;
  - Détermination des actions en fonction des parties du corps;
  - Détermination de la pose (plus ou moins précise);
  - Reconnaissance de gestes.

# Modélisation d'humains

---

## □ Par la silhouette

### ■ Avantages:

- Silhouette facile à obtenir;
- Insensible aux couleurs et textures;
- Donne des informations partielles sur la pose.

### ■ Inconvénients:

- Ombres déforment la silhouette;
- Ambiguïtés sur la pose (dos/face ?);
- Occlusions internes (bras à l'intérieur de la région du corps).

# Modélisation d'humains

---

## □ Par la silhouette

- Exemple: avec shape context.
- Utilise un descripteur local avec des points équidistants sur la silhouette. (400-500 points pour une silhouette de 64 x 128 pixels).
- Histogramme du nombre de pixels frontières (sur la silhouette) dans un référentiel log-polaire de rayon prédéterminé.

# Modélisation d'humains

---





# Modélisation d'humains

---

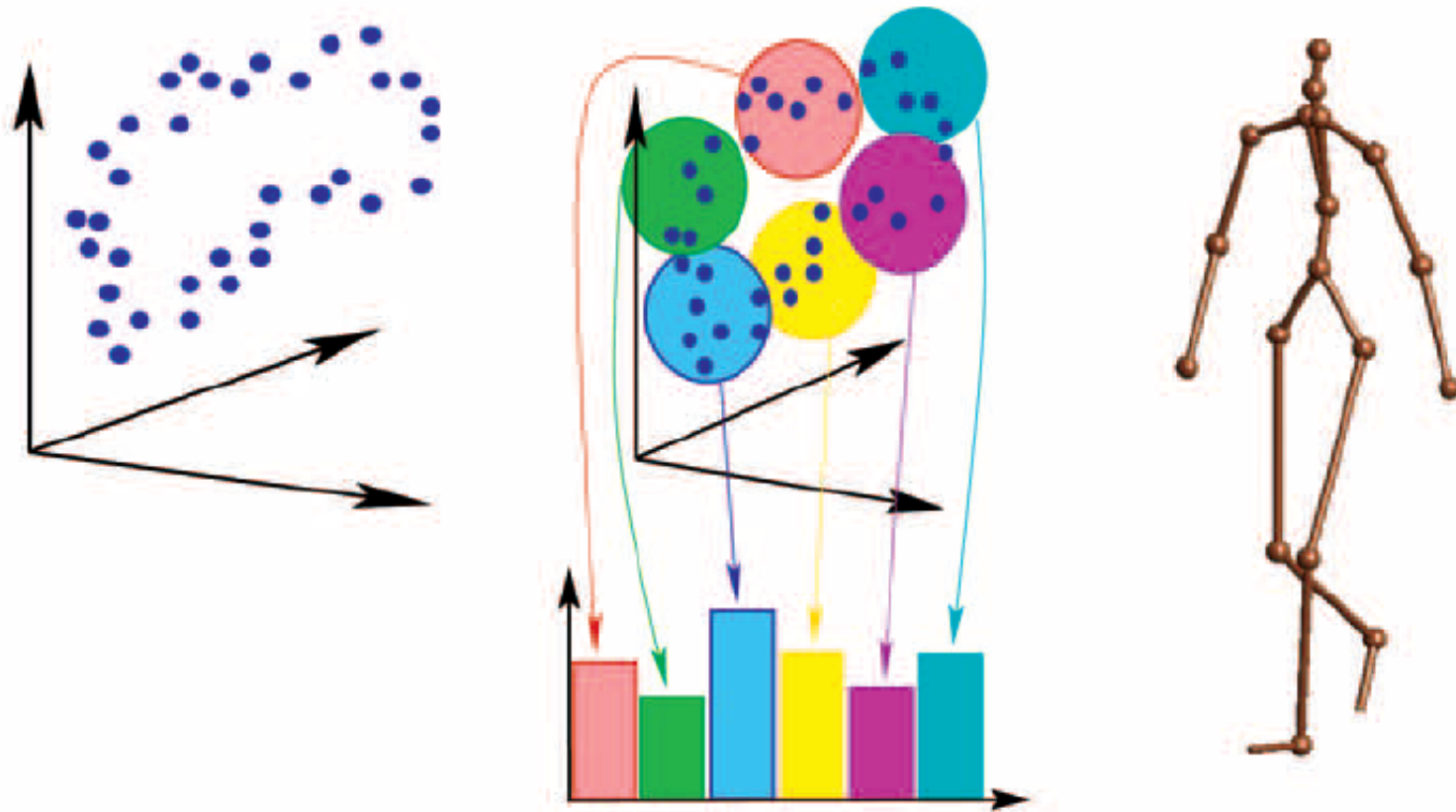
- ❑ Par la silhouette (*suite*)
  - Ensuite, tous les histogrammes sont exprimés sous forme de vecteurs (angle, distance, nombre de données).
  - Les vecteurs sont groupés en 100 classes (par *clustering*).
  - On calcule un histogramme avec 100 classes, dont la fréquence de chaque classe est le nombre de vecteurs dans chacune de celles-ci.

# Modélisation d'humains

---

- ❑ Par la silhouette (*suite*)
  - À chaque nouvelle silhouette, on construit cet histogramme en fonction des 100 classes apprises.
  - En construisant un modèle par pose, on peut obtenir la pose d'une silhouette trouvée dans une séquence vidéo en comparant l'histogramme obtenu avec l'histogramme-modèle.

# Modélisation d'humains



# Modélisation d'humains

---

- ❑ Par la silhouette
  - Utilise trois modèles avec des fonctions différentes.
    - ❑ Histogrammes de projection;
    - ❑ Patrons temporels de textures;
    - ❑ Localisation des parties du corps.

# Modélisation d'humains

---

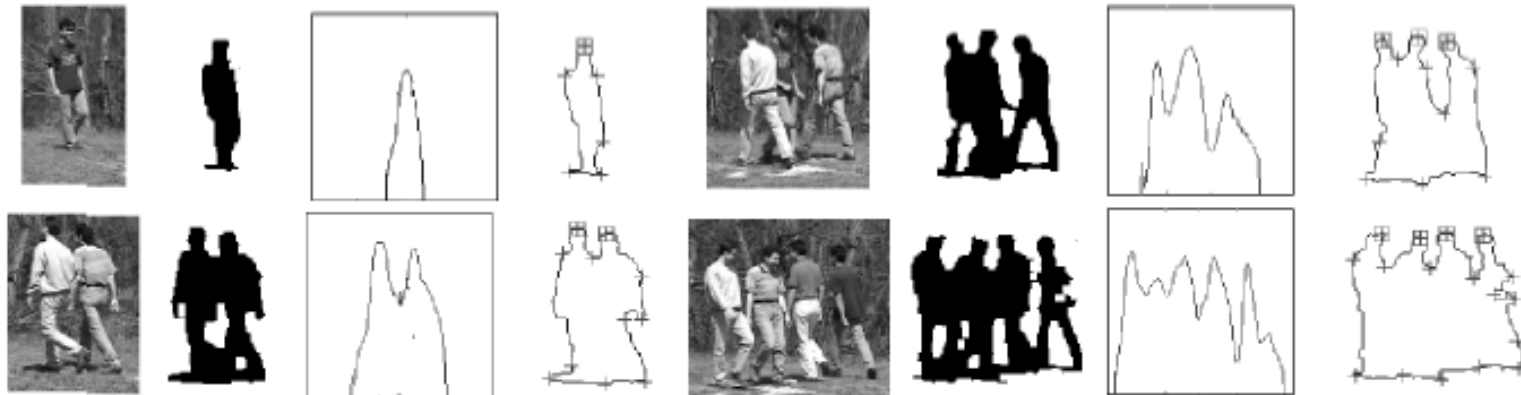
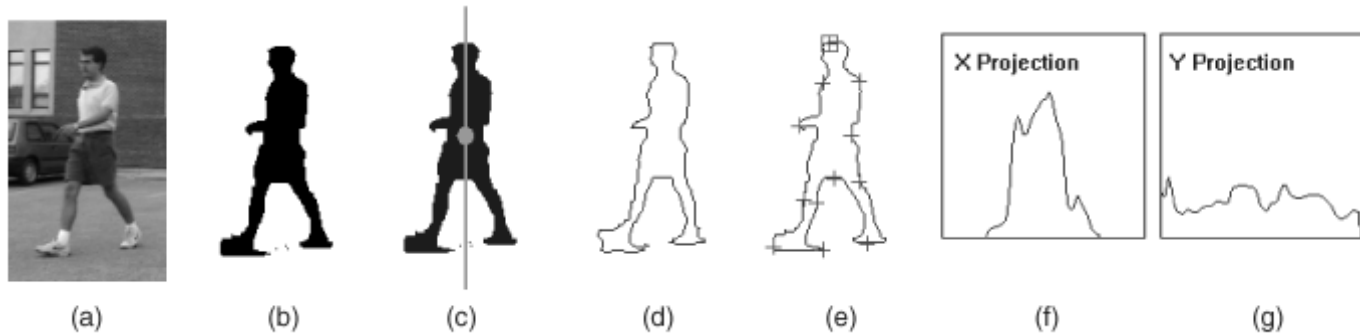
- ❑ Par la silhouette *(suite)*
  - Histogrammes de projection
    - ❑ Localisation de l'axe majeur par analyse en composantes principales (PCA). Vecteur propre où la covariance est la plus élevée pour les positions x et y du blob.
    - ❑ C'est-à-dire: On trouve l'axe passant par le centroïde qui minimise la somme des distances perpendiculaires par rapport à l'axe.
  - Exemple MATLAB: *PCAaxeprincipal.m*

# Modélisation d'humains

---

- ❑ Par la silhouette *(suite)*
  - Histogrammes de projection
    - ❑ Prenant l'axe majeur comme référence, on calcule les histogrammes du nombre de pixels projetés à chaque position en X et en Y.
    - ❑ Permet de différencier un humain par rapport à d'autres types d'objets. Permet aussi de détecter le nombre de personnes dans un attroupement.

# Modélisation d'humains



# Modélisation d'humains

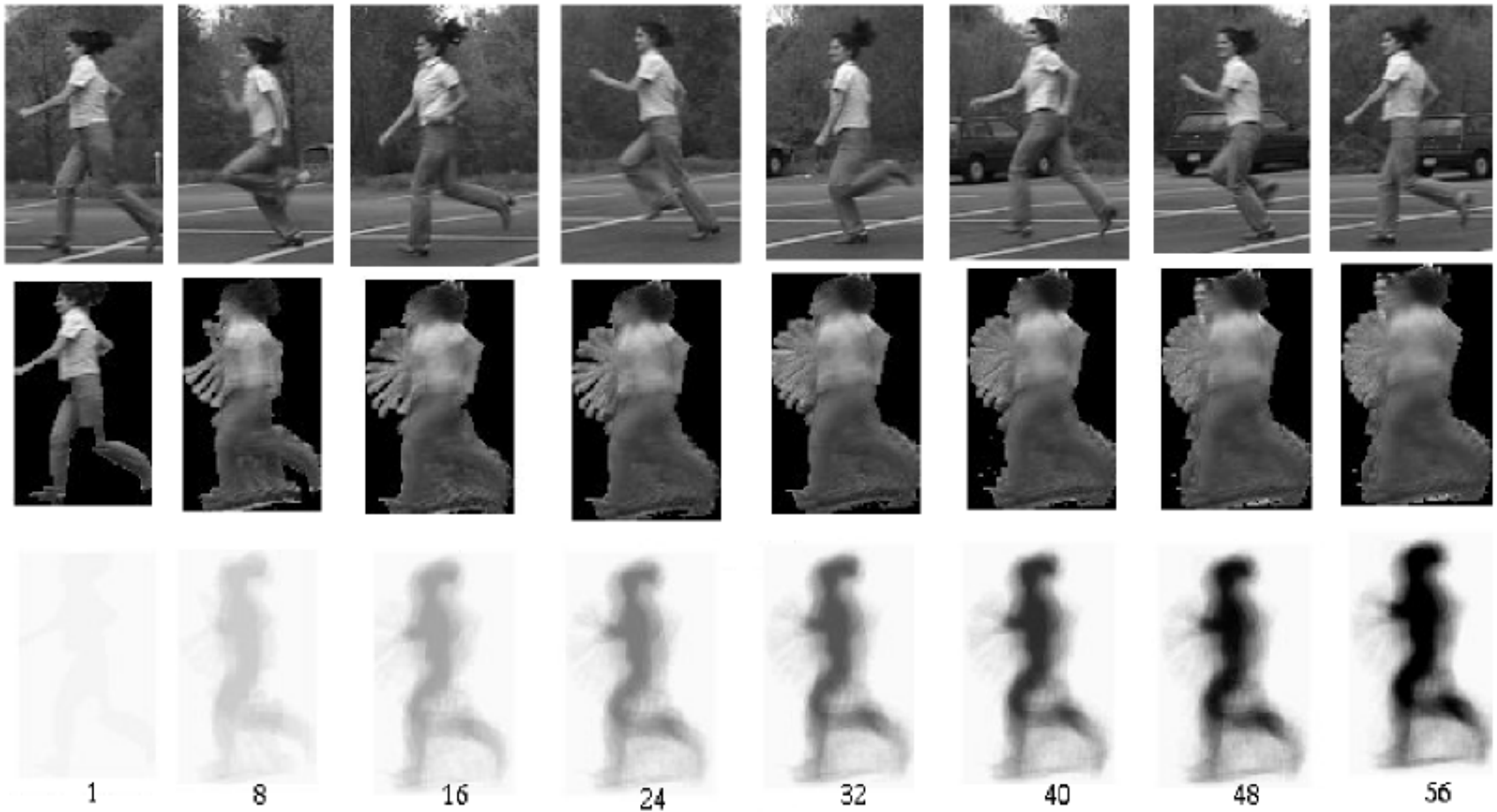
---

- Par la silhouette *(suite)*
  - Patrons temporels de textures
    - Modélisation d'une personne par un modèle d'apparence dynamique.
    - Somme des intensités des pixels  $I(x,y)$  pondérée par le nombre de fois qu'ils sont détectés comme objet sur  $N$  trames ( $w^{t-1}(x,y)$ ):

$$\Psi^t(x, y) = \frac{I(x, y) + w^{t-1}(x, y) \times \Psi^{t-1}(x, y)}{w^{t-1}(x, y) + 1}$$



# Modélisation d'humains



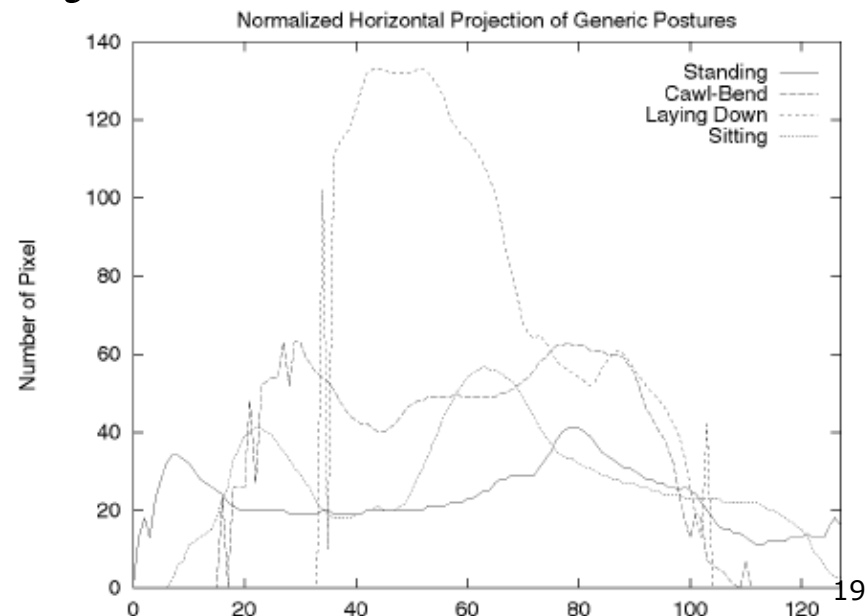
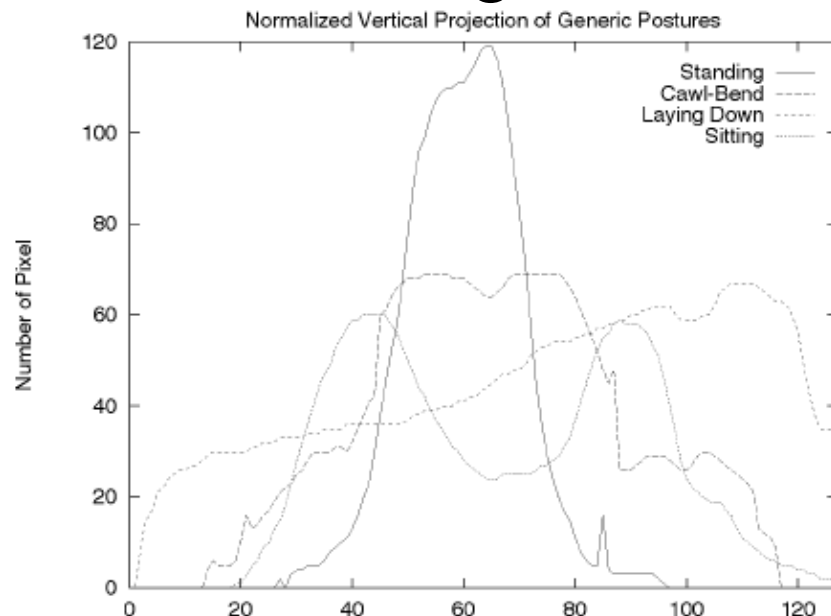
# Modélisation d'humains

---

- ❑ Par la silhouette (*suite*)
  - Patrons temporels de textures
    - ❑ Permet de mettre en correspondance une personne lorsqu'elle apparaît/disparaît due à une occlusion.
    - ❑  $w^t(x,y)$ : *Modèle de forme*
    - ❑  $\Psi^t(x,y)$ : *Texture en tons de gris*
    - ❑ Compare par corrélation.

# Modélisation d'humains

- Par la silhouette (*suite*)
  - Localisation des parties du corps
    - Détermination de la posture par les histogrammes de projection.

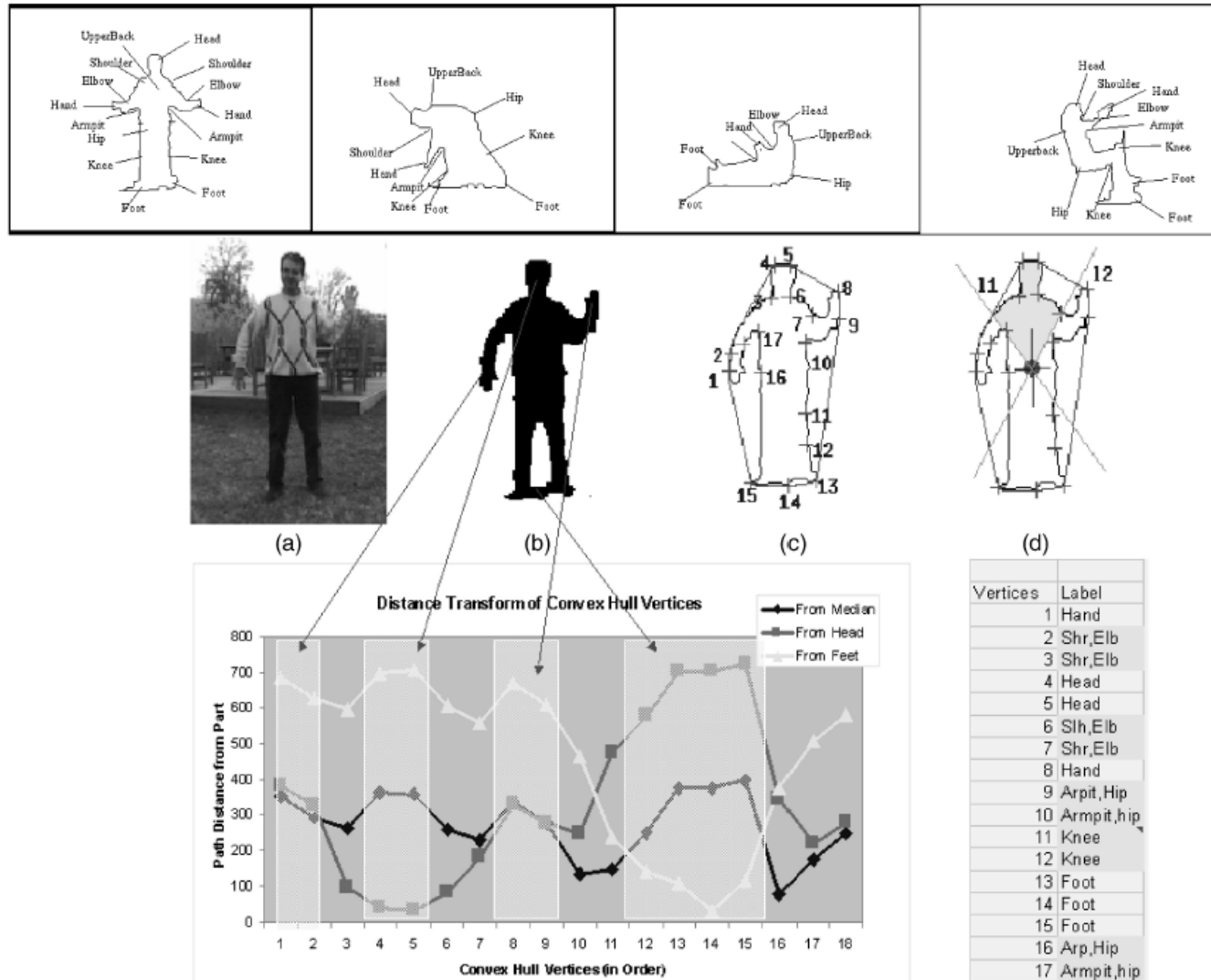


# Modélisation d'humains

---

- ❑ Par la silhouette *(suite)*
  - Localisation des parties du corps
    - ❑ Détection des points de convexité.
    - ❑ Tête est au haut de l'axe majeur. Ensuite, les autres parties du corps apparaîtront sur la silhouette dans un ordre dépendant de la posture et basé sur la transformée de distance.

# Modélisation d'humains



# Modélisation d'humains

---

## □ Par régions

### ■ Avantages:

- Permet d'obtenir les parties du corps.
- Donne des informations plus précises sur la pose.
- Occlusions internes peuvent être gérées.

### ■ Inconvénients:

- Plus difficile à obtenir de façon fiable (Sensible aux couleurs et textures);
- Ambiguïtés sur la pose (dos/face ?);

# Modélisation d'humains

---

## □ Par régions

- Utilise un modèle basé sur les régions et leurs positions.
- Après soustraction d'arrière-plan et détermination des composantes connectées, on caractérise des segments par plusieurs propriétés unaires, binaires et tertiaires.
- Pour trouver les segments, on segmente les régions connectées en sous-régions à peu près uniformes. On suppose qu'à cause des occlusions, une région connectée peut contenir deux objets.

# Modélisation d'humains

---

## □ Par régions *(suite)*

### ■ Unaire:

- Taille
- Couleur (moyenne de H,S,V)
- Localisation (centroïde)
- Périmètre (voisinage de 8)

### ■ Binaire:

- Ratio de frontière (Nombre de pixels de contour de **A** qui touchent le contour de **B** divisé par le nombre de pixels de contour de **A**).
- Adjacence



# Modélisation d'humains

---

## □ Par régions *(suite)*

### ■ Tertiaire:

#### □ Adjacence à trois

- Pour réduire les erreurs, on fait d'abord une fusion des segments à l'aide des propriétés unaires, binaires, et tertiaires + dimension.
- (Pour le suivi, seules les propriétés unaires sont utilisées.)

# Modélisation d'humains

A



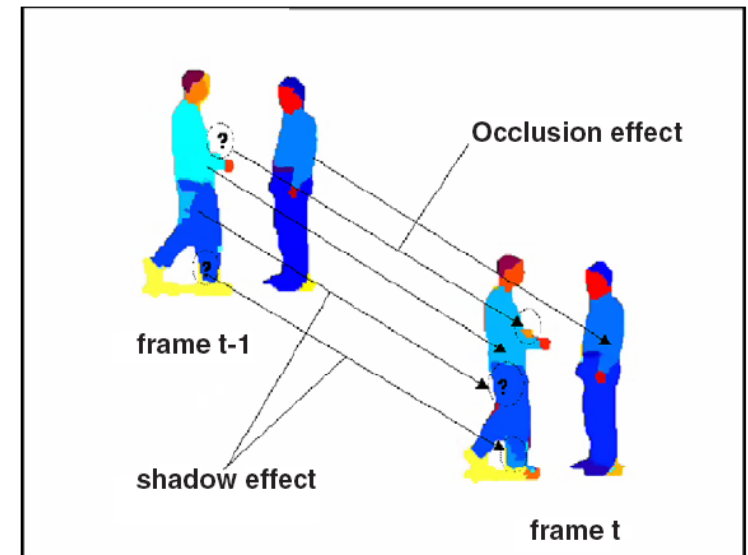
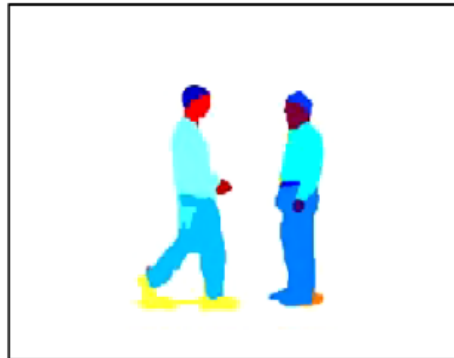
B



A



B



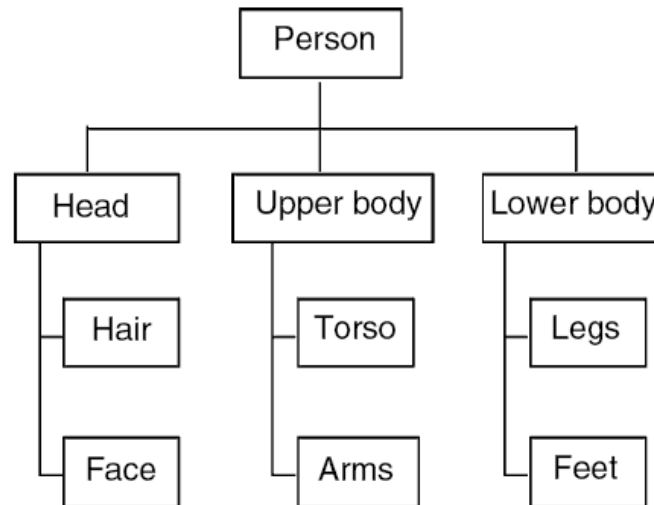
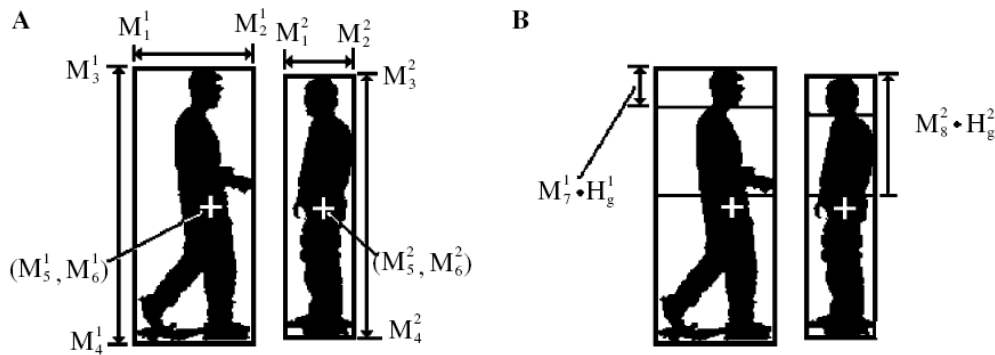
# Modélisation d'humains

---

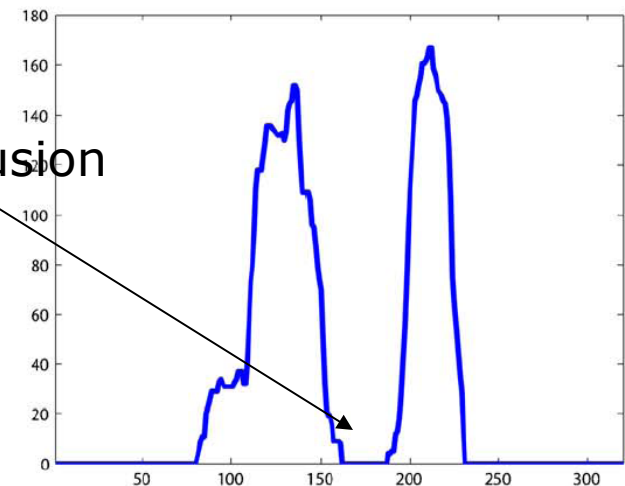
## □ Par régions *(suite)*

- On positionne les segments à l'intérieur du rectangle qui les englobe. On initialise le système lorsqu'il n'y a pas d'occlusions, donc groupes de segments=composantes connectées.
- Si groupe de segments forme un humain, leurs tailles respecteront les régions têtes/corps/jambes établies selon des études anthropométriques.
- Finalement, on étiquette aussi chaque segment comme étant une région de peau ou non.

# Modélisation d'humains



Pas  
d'occlusion



# Modélisation d'humains

---

## ❑ Par bâtonnets

### ■ Avantages:

- ❑ Permet d'obtenir les parties du corps et leurs relations.
- ❑ Donne des informations précises sur la pose.

### ■ Inconvénients:

- ❑ Obtention dépend de régions ou du contour.
- ❑ Ambiguïtés sur la pose (dos/face ?);

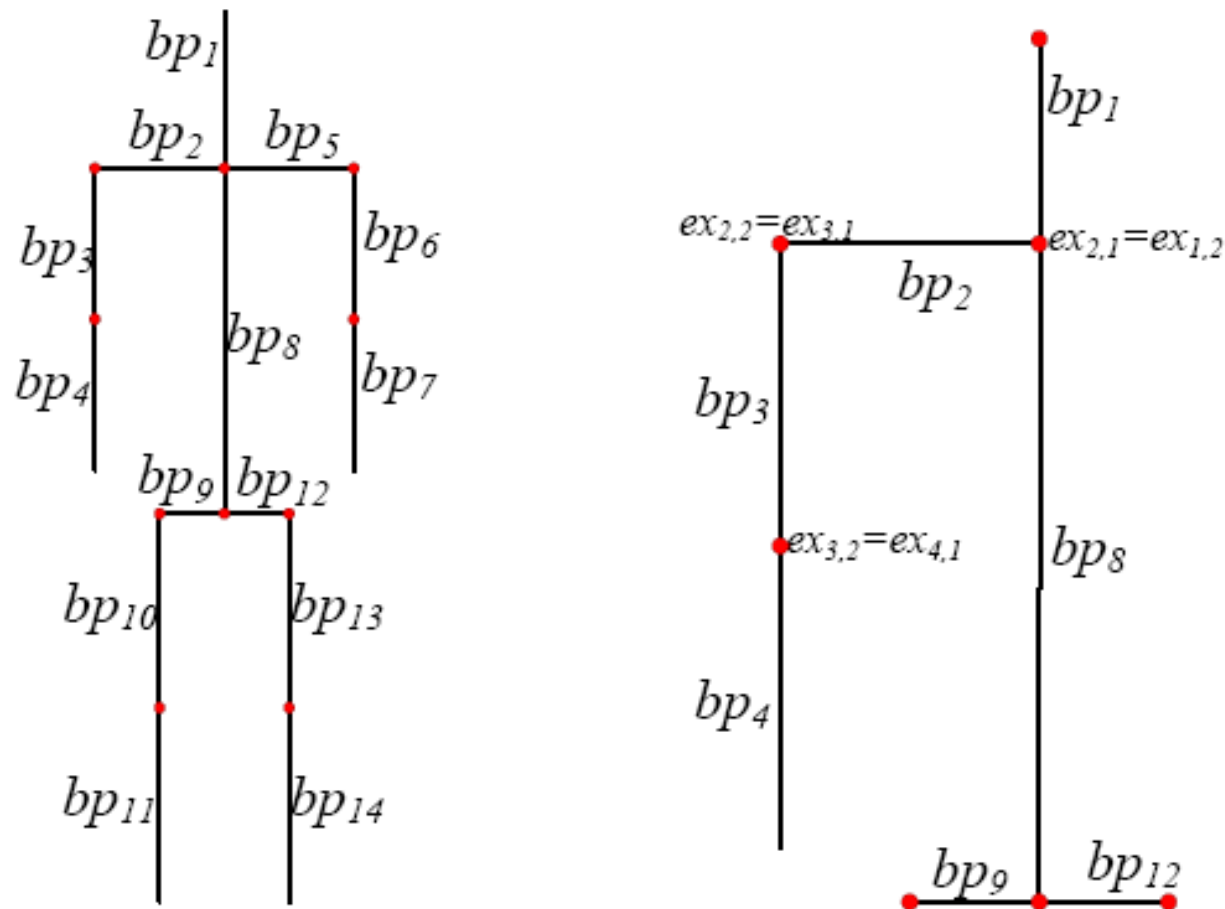
# Modélisation d'humains

---

## □ Par bâtonnets

- Utilise l'image binaire d'un humain, et y ajuste un modèle par bâtonnets.
- Les parties du corps sont définies par un vecteur de 14 segments de droite représentant la tête, le tronc et les membres.
- Les proportions des segments sont basées sur des études anthropométriques.
- Entre chaque segment, des contraintes sont définies sur les angles possibles pour les connexions.

# Modélisation d'humains



# Modélisation d'humains

---

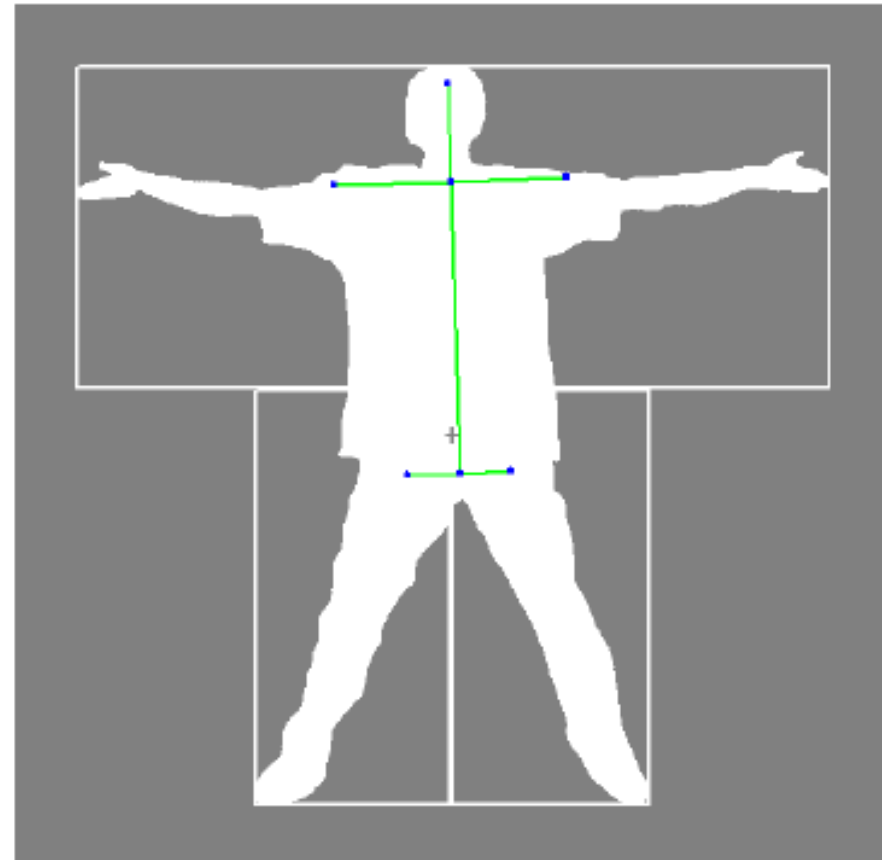
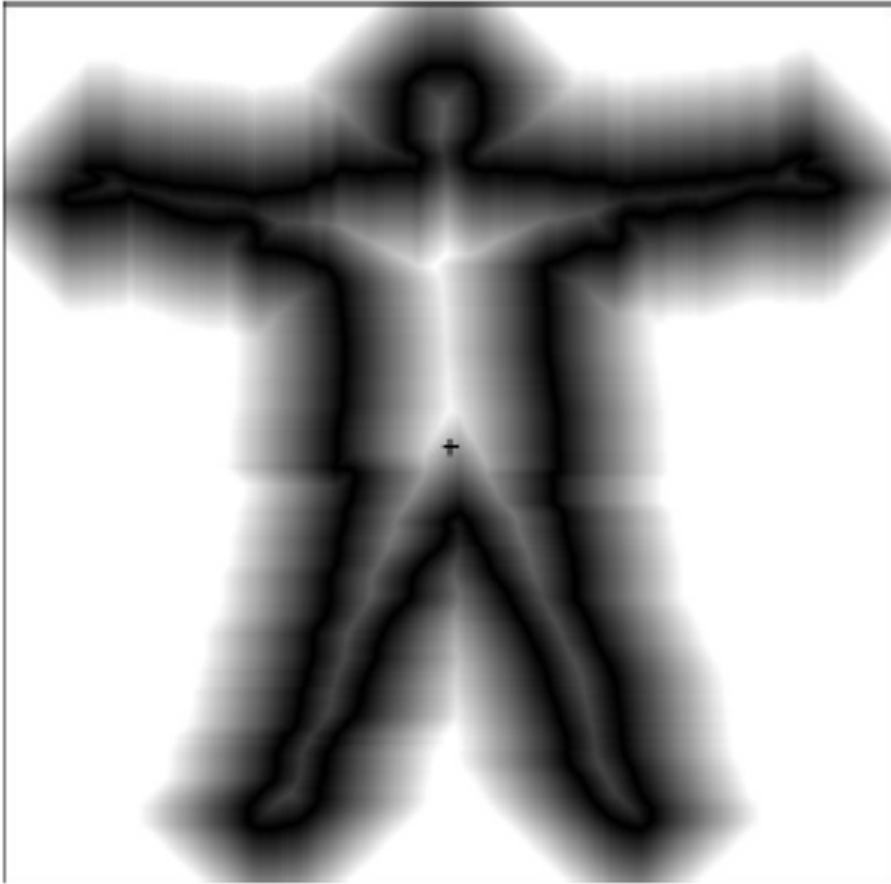
## □ Par bâtonnets *(suite)*

- Le tronc est ensuite placé en premier à partir de proportion (tête à  $2/15$ , hanches à  $8/15$  du corps) sur l'axe principal du corps.
- L'axe principal du corps est trouvé par la transformée de distance.
- Transformée de distance: distance de chaque pixel par rapport à la frontière de la région.
- Choix du modèle frontal ou latéral en fonction de la largeur de la région perpendiculairement à l'axe principal.



# Modélisation d'humains

---



# Modélisation d'humains

---

## □ Par bâtonnets *(suite)*

- Pour les membres, ils sont placés un par un de façon progressive.
- On génère toutes les possibilités respectant les contraintes d'angle des articulations, et sélectionne le meilleur segment par:
  - est le plus possible à l'intérieur de la région du corps (transformée de distance);
  - permet de s'approcher le plus de la frontière extérieure du rectangle englobant pour le membre.

# Modélisation d'humains

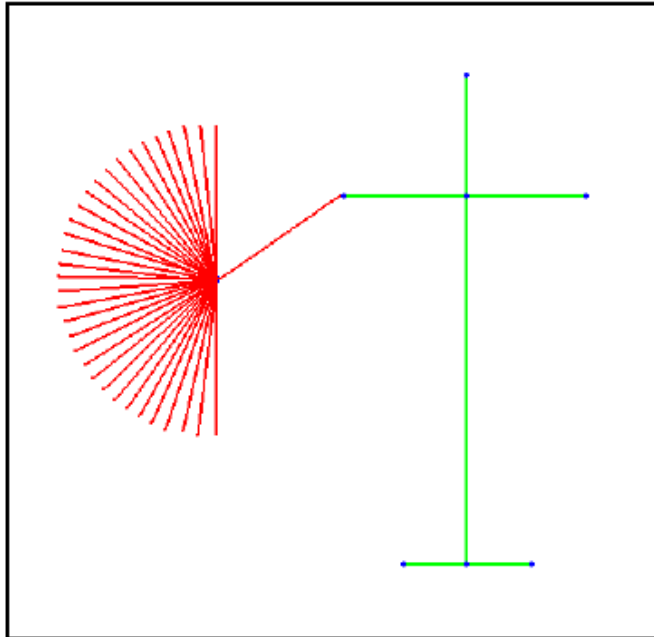
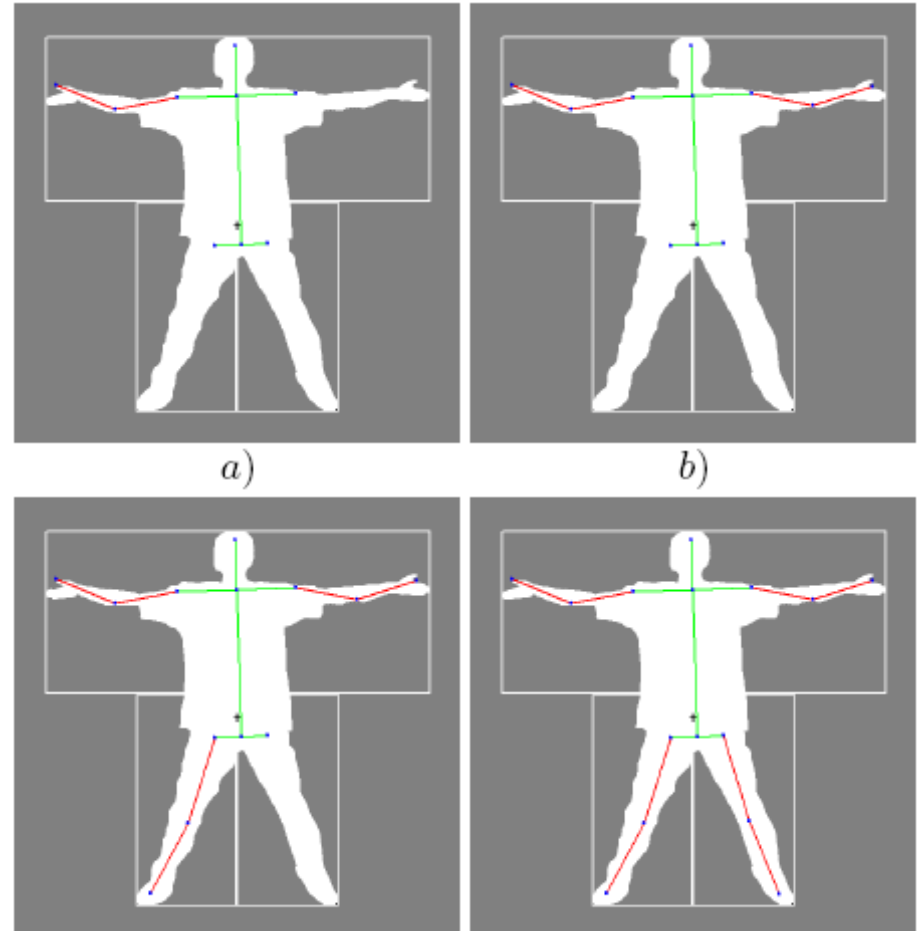


Figure 6: One particular position for  $bp_3$  and all the candidate positions generated by the system for  $bp_4$ . This process is repeated for all possible positions of  $bp_3$  admitted by the angle constraints. The sampling angle in this case is  $\pi/32$ .



# Modélisation d'humains

---

## ❑ Par régions et bâtonnets

- On segmente l'image en régions (pas avec des séquences vidéo);
- On identifie des parties de bras et de jambes.
- Quatre mesures sur les régions:
  - ❑ Saillance du contour (Probabilité que la frontière d'une région corresponde à une vraie frontière entre deux objets)
  - ❑ Recouvrement entre un rectangle mis à l'échelle (approximation d'un membre) avec la région.

# Modélisation d'humains

---



# Modélisation d'humains

---

- ❑ Par régions et bâtonnets *(suite)*
  - ❑ Effets d'ombre sur les membres.
  - ❑ Focus de la région. Région des membres moins floue (si suppose humain au focus).
- Après, trouve les membres complets en comparant les régions aux alentours. Un membre complet sera composé des régions les plus similaires parmi celles aux alentours.
- La dimension des membres donne l'échelle de la personne.

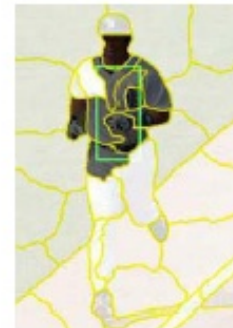
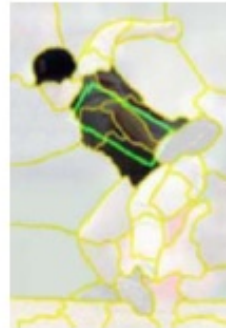
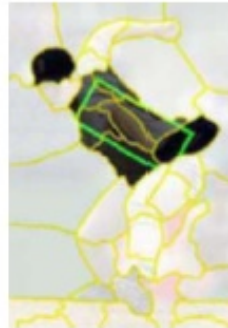
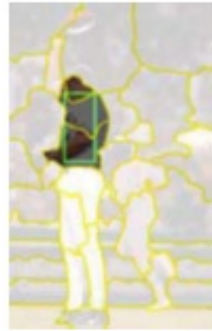
# Modélisation d'humains

---

- ❑ Par régions et bâtonnets *(suite)*
  - Pour trouver le torse, on teste toutes les combinaisons de régions qui entrent à l'intérieur d'un rectangle englobant (selon l'échelle des membres).
  - Utilise les mesures de contour, forme et focus.
  - Cherche la tête de la même façon avec un disque comme modèle mis à l'échelle par les proportions du torse. Tient compte de la position relative entre la tête et le torse.

# Modélisation d'humains

---





# Modélisation d'humains

---

## □ Par régions et bâtonnets *(suite)*

- On assemble le tout, en sélectionnant les meilleurs candidats (c'est-à-dire que le membre avec la meilleure mesure n'est pas nécessairement sélectionné. C'est la mesure globale pour toute la configuration de membres/torse/tête qui compte).
- Sélection avec des contraintes sur:
  - Dimension relative
  - Symétrie
  - Adjacence

afin de n'étudier que les configurations les plus plausibles.

# Modélisation d'humains



# Reconnaissance de démarches

---

## □ Reconnaissance de démarches

- Le nombre de pixels de l'image binaire de l'humain permet de distinguer les démarches.
- L'humain est détecté par soustraction d'arrière-plan. On extrait alors l'image binaire du marcheur.
- La région binaire de l'humain est mise à l'échelle de façon à ce que sa hauteur soit de 128 pixels.

# Reconnaissance de démarches

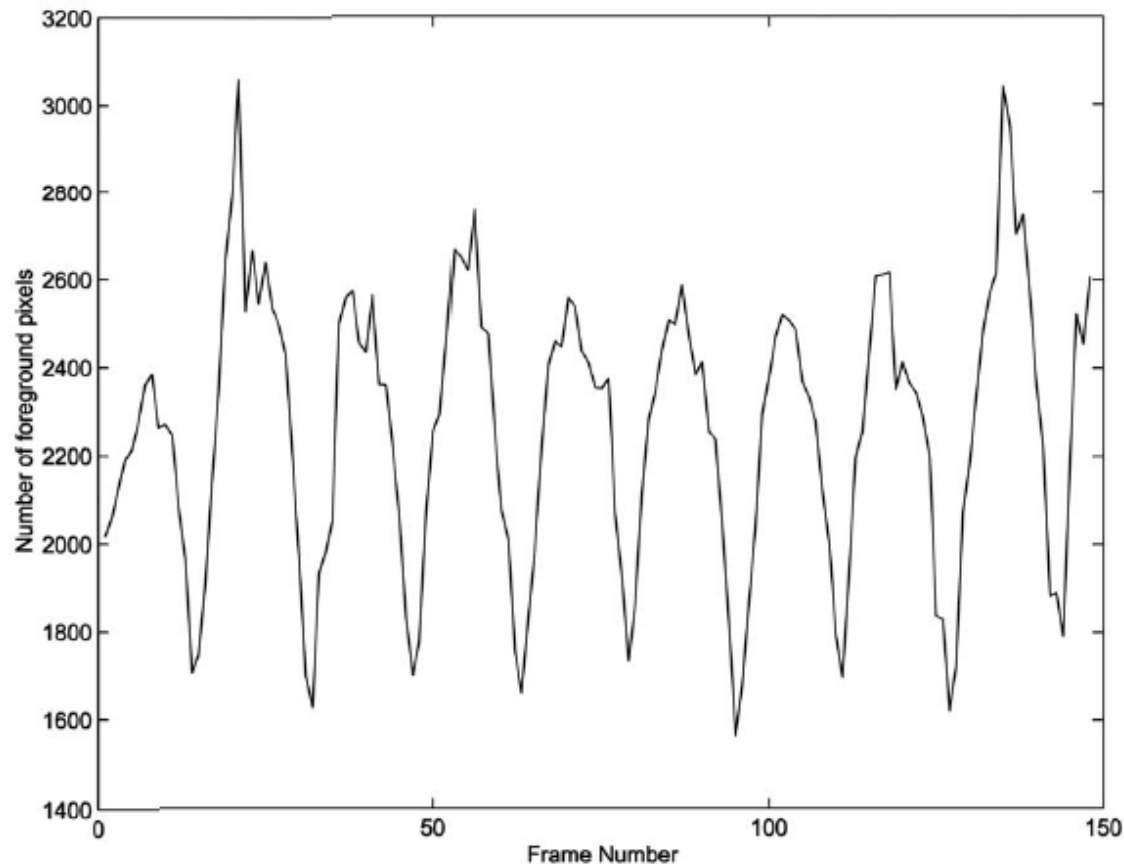
---

## □ Reconnaissance de démarches *(suite)*

- Le principe de la méthode est que le nombre de pixels de la région du marcheur varie en fonction de sa pose. S'applique seulement à des vues latérales.
- Le nombre de pixels de la région est plus grand lorsque les jambes sont bien séparées, alors qu'il est plus petit lorsque les jambes sont en recouvrement.
- Pour une plus grande sensibilité, le calcul est fait avec la moitié inférieure de la région.

# Reconnaissance de démarches

## □ Reconnaissance de démarches (*suite*)



# Reconnaissance de démarches

---

## □ Reconnaissance de démarches (*suite*)

- La période est évaluée par:
  - La médiane de la distance entre chaque 2 minimum en partant du premier minimum;
  - La médiane de la distance entre chaque 2 minimum en ignorant le premier minimum.
- La période  $N_{gait}$  est la moyenne des médianes.
- Les démarches sont comparées sur  $N_{gait}$  images en comparant le ratio de intersection sur l'union des pixels de la région du marcheur.

# Reconnaissance de démarches

---

## □ Reconnaissance de démarches (*suite*)

- C'est-à-dire, qu'on calcule la corrélation entre les régions des marcheurs  $S_P$  et  $S_G$ :

$$\text{Corr}(S_{Pk}, S_G)(l) = \sum_{j=0}^{N_{\text{exit}}-1} \text{FrameSim}(S_P(k+j), S_G(l+j))$$

- où

$$\text{FrameSim}(S_P(i), S_G(j)) = \frac{\text{Num}(S_P(i) \cap S_G(j))}{\text{Num}(S_P(i) \cup S_G(j))}$$

# Reconnaissance de démarches

## □ Reconnaissance de démarches (*suite*)

TABLE 5

Baseline Performances for the Challenge Experiments in Terms of the Identification Rate  $P_I$  at Ranks 1 and 5 and the Verification Rate  $P_V$  at a False Alarm Rate of 1 Percent and 10 Percent of Unnormalized (UN), Z-Norm (ZN), and MAD-Norm (MAD)

Exp.	Difference	$P_I$ (%) (at rank)		$P_V$ (%) at $P_F = 1\%$			$P_V$ (%) at $P_F = 10\%$		
		1	5	UN	MAD	ZN	UN	MAD	ZN
A	View	73	88	52	80	82	81	94	94
B	Shoe	78	93	48	80	87	82	94	94
C	Shoe, View	48	78	32	57	65	69	89	94
D	Surface	32	66	24	36	44	61	80	80
E	Surface, Shoe	22	55	16	33	35	52	76	76
F	Surface, View	17	42	10	22	20	45	59	60
G	Surface, Shoe, View	17	38	12	24	28	40	57	55
H	Briefcase	61	85	46	68	72	80	90	91
I	Briefcase, Shoe	57	78	48	60	67	76	85	85
J	Briefcase, View	36	62	22	45	48	64	75	76
K	Time, Shoe, Clothes	3	12	0	3	6	15	27	24
L	Surface, Time, Shoe, Clothes	3	15	0	3	6	18	27	24

All performance scores are in percent.



# Bibliographie

---

- ❑ A.Agarwal,B.Triggs, Recovering 3D human Pose from monocular images, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 28, No.1, 2006, pp. 44-58
- ❑ A.Bobick et al., The Kidsroom: A Perceptually-Based Interactive and Immersive Story Environment, MIT Media Laboratory, Technical Report No. 398, 1997
- ❑ I. Haritaoglu et al. W4: Real-time Surveillance of People and Their Activities, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 22, No.8, 2000, pp. 809-830
- ❑ G.Mori et al., Recovering Human Body Configurations: Combining Segmentation and Recognition, in CVPR'04, 2004
- ❑ S.Park, J.K. Aggarwal, Simultaneous tracking of multiple body parts of interacting persons, Computer Vision and Image understanding, 102, 2006, pp. 1-21
- ❑ K.Sato,J.K. Aggarwal, Temporal spatio-velocity transform and its application to tracking and interaction, Computer Vision and Image Understanding 96, 2004, pp. 100–128
- ❑ J.Vignola et al., Progressive Human Skeleton Fitting, in Vision interface 2003, 2003, pp 35-42
- ❑ S. Sarkar et al., The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 2, 2005, pp. 162-177