

# Segmentation couleur et condensation pour le suivi et la reconnaissance de gestes humains

L.Brèthes<sup>†</sup>, P.Menezes<sup>‡</sup>, F.Lerasle<sup>†</sup>, M.Briot<sup>†</sup>

<sup>†</sup> LAAS-CNRS, 7 avenue Colonel Roche, 31077 Toulouse Cédex 4

<sup>‡</sup> ISR/DEEC, Pinhal de Marrocos, Université de Coimbra, 3030-290 Coimbra

e-mail: {lbrethes, lerasle, briot}@laas.fr, pm@deec.uc.pt

## Résumé

*L'interaction homme/machine est une problématique majeure dans la communauté Vision pour la Robotique, notamment dans la perspective d'un robot personnel. Un des enjeux dans ce contexte est de développer un robot capable de détecter et de suivre des gestes effectués par une personne à proximité. Cet article présente une approche basée sur du filtrage particulière permettant de réaliser ces fonctions visuelles pour une main, éventuellement une tête (pour le suivi de simples usagers de l'environnement). Dans un contexte d'interaction active, le but est aussi de reconnaître quelques configurations canoniques de la main. Chaque configuration est modélisée par une spline représentant sa forme. Sa position est estimée en recalant la forme prototype (template) sur l'image du gradient couleur filtré par le résultat d'une segmentation couleur. L'originalité de la méthode de segmentation réside dans l'application séquentielle d'un algorithme de ligne de partage des eaux sur la chrominance puis la luminance des pixels de la peau. La reconnaissance du geste courant et la commutation éventuelle entre templates sont réalisées dans la boucle de suivi. Des résultats sont ensuite illustrés et montrent notamment la robustesse de la méthode aux environnements complexes et aux conditions d'illuminations changeantes. Les limites ainsi que les extensions prévues sont discutées en final.*

## Mots-Clef

vision couleur, appariement de forme prototype, condensation, suivi, reconnaissance

## Abstract

*The interaction between man and machines has become an important topic for the robotic community as it can generalise the use of robots. One of the requirements for this interaction is that a robot be able to detect and track hand gesture performed by a person in its vicinity. This paper describes a particle filter based method enabling the achie-*

*vement of these visual functions for human hand (or human head when the aim is to track environment users). In an active interaction context, we want to recognize some canonical hand gestures. Each hand gesture is modelled by a 2D spline. Its pose is estimated by fitting the model onto the current image colour gradient filtered by a colour segmentation. The originality of our segmentation method consists to apply sequentially a watershed algorithm based on chrominacity then intensity on skin colour pixels. The recognition of the current hand gesture and commutations between associated templates are achieved in the tracking loop. Results of tracking and recognition using both shape and color cues are illustrated in the paper and show the process robustness in cluttered environments and in various light conditions. The limits of the method and future works are also discussed.*

## Keywords

colour vision, template matching, watershed, condensation, tracking, recognition.

## 1 Introduction

Durant ces dernières années, beaucoup de travaux dans la communauté Vision pour la Robotique ont porté sur l'interaction entre l'homme et la machine ([8],[15]). Un défi majeur de la Robotique est sans doute celui du robot personnel avec notamment la perspective de permettre à un robot mobile de naviguer de manière autonome en présence de public.

Plus qu'un simple usager de l'environnement qu'il partage avec les humains, le robot doit en outre pouvoir interagir avec ces derniers par la reconnaissance de gestes élémentaires pour l'exécution de tâches décidées par eux : asservissement sur leurs déplacements, apprentissage supervisé, manipulation d'objets,... Dans ce contexte, les travaux présentés dans cet article portent plus spécifiquement sur la détection, le suivi et la reconnaissance de gestes élémentaires à partir du flot vidéo d'une caméra couleur.

Pour lever l'ambiguïté mouvement-profondeur, certaines approches exploitent des modèles volumiques issus de connaissances géométriques et cinématiques sur les membres corporels observés. Ces modèles sont souvent articulés ([16],[7],[18]), voire déformables ([11],[19]). Pour toutes ces approches, la modélisation et l'évaluation de tous les paramètres de déformations (même par des techniques récentes d'estimation) restent lourdes dans leurs mises en œuvre.

De notre point de vue, les approches basées sur des modèles 2D offrent un meilleur compromis entre temps de traitement et pertinence de l'analyse. Parmi toutes les approches 2D développées, nous pouvons citer les travaux s'appuyant sur les contours actifs, notamment les *snakes* associés à du filtrage particulier pour Isard *et al.* [10] ou les EDP pour Paragios *et al.* [14]. Ces techniques permettent de prendre en compte les déformations d'une forme prototype (*template*). Des variantes consistent à définir un modèle statistique hiérarchique [12] voire à gérer plusieurs *templates* [9] pour gérer les déformations. En plus de la forme, certains *templates* agrègent des informations de texture ou de couleur [6] afin de robustifier le processus de suivi.

Les méthodes d'analyse estimant à la fois le mouvement et la structure des membres nécessitent de résoudre un problème complexe et conduisent parfois à des solutions instables en présence de bruit. Des perturbations sont cependant incontournables lorsqu'aucune restriction/hypothèse n'est faite quant au contexte de la prise d'images.

La prise en compte de ces considérations et la volonté de simplifier au maximum la modélisation font que nous avons privilégié une approche ne requérant qu'une forme prototype grossière non déformable de la silhouette de la cible à suivre. La section 2 décrit brièvement notre approche tandis que les sections 3 et 4 se focalisent respectivement sur la segmentation couleur et le processus d'estimation pour le suivi et la reconnaissance de configurations de la main. Des résultats de suivi et de reconnaissance sont décrits et discutés dans la section 5. La section 6 propose enfin quelques extensions envisagées pour ces travaux.

## 2 Description de l'approche

La forme prototype représente ici une main pour une interaction active, éventuellement une tête dans un contexte de suivi de personnes (interaction passive ou active). Les contours de cette forme sont modélisés par une spline (figure 1) dont l'état est estimé par un filtre particulier.

Contrairement à Blake *et al.* qui utilisent la couleur de façon marginale dans leurs travaux, une segmentation en régions sur la base d'attributs colorimétriques, décrite dans la section suivante, est partie intégrante du processus de suivi. La combinaison d'informations hétérogènes telles que la forme et la couleur permet de robustifier le processus de suivi dans notre contexte applicatif. En effet, les prise

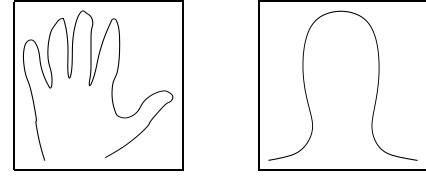


FIG. 1 – Modèles de silhouettes de main ou tête

de vues s'effectuant depuis une caméra embarquée sur un robot mobile, la complexité de la scène (son arrière plan notamment) et les conditions d'éclairage (*a priori* quelconques) peuvent être génératrices de fausses mesures.

A notre connaissance, peu de travaux réalisent le suivi et la reconnaissance simultanément. Dans ce cadre, citons les travaux de Bretzner *et al.* [5] qui définissent un modèle hiérarchique grossier de la main constitué de cercles et d'ellipses. Cette démarche permet d'espérer des temps de traitement compatibles avec le contexte applicatif qui est le nôtre. Chaque configuration de la main à reconnaître est ici caractérisée par un *template* (figure 2). La reconnaissance et la commutation d'un *template* à l'autre s'effectuent dans le processus d'estimation dédié au suivi.

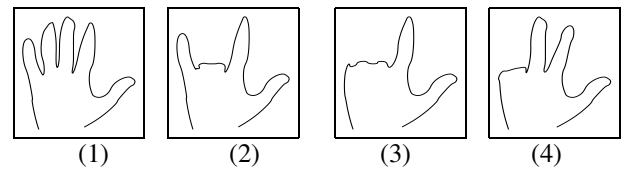


FIG. 2 – Les quatre configurations à reconnaître

## 3 Segmentation couleur

### 3.1 Principe

La méthode de segmentation couleur développée combine une phase de détection des pixels de la peau et une phase de *clustering* par application successives d'un algorithme de lignes de partage des eaux (LPE) [3]. Une segmentation non supervisée en régions correspondant potentiellement à des membres corporels (main, tête,...) est ainsi réalisée. Cette deuxième phase a pour but de regrouper les pixels connexes qui sont cohérents d'un point de vue chrominance et luminance.

En raisonnant sur un algorithme de LPE, notre approche s'appuie donc sur la morphologie de l'histogramme et se démarque ainsi des approches purement statistiques qui reposent sur des modèles (multi)-gaussiens souvent éloignés de la réalité. De plus, l'algorithme de LPE calcule automatiquement le nombre de classes à partir d'un nombre de marqueurs pré-définis. Le choix de ces marqueurs est donc essentiel. Un algorithme de LPE offre enfin des temps de calcul compatibles avec les applications visées.

La figure 3 résume notre approche. Les différentes étapes sont détaillées ci-après.

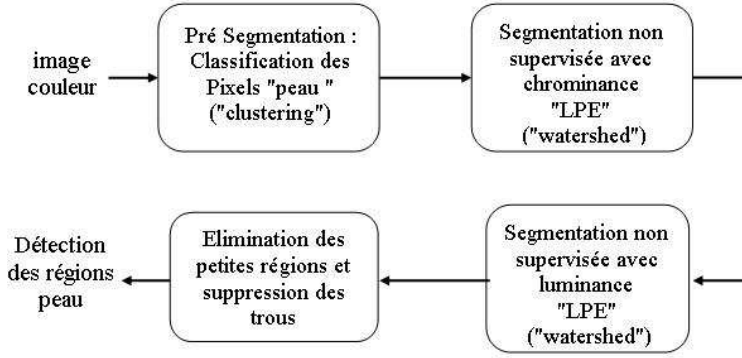


FIG. 3 – Algorithme de segmentation couleur

### 3.2 Classification des pixels

Différentes bases colorimétriques ont été considérées [2]. Nous avons finalement opté pour la base colorimétrique  $(I_1, I_2, I_3)$  en raison de ses performances connues en terme de séparabilité des classes [13]. Étant basée sur un apprentissage préalable des classes peau et non peau, cette étape considère uniquement la composante chrominance  $(I_2, I_3)$  afin de limiter l'influence des conditions d'éclairage ambiantes lors des prises de vue. Nous avons choisi le rapport d'histogrammes [17] pour classer les pixels car aucune hypothèse limitatrice n'est faite quant à la distribution (gaussienne par exemple) des composantes  $I_2$  et  $I_3$  sur chacune des deux classes. Les histogrammes de l'arrière-plan  $h_{non-peau}$  et de la peau  $h_{peau}$  sont élaborés par apprentissage sur une base d'images représentatives des scènes observées par le robot au cours de sa navigation dans l'environnement (défini par un réseau de couloirs et d'espaces ouverts).

La probabilité  $p(peau|C)$  d'appartenance à la classe peau d'un pixel de couleur  $C$  est donnée classiquement par la règle de Bayes, soit après simplification :

$$p(peau|C) = \frac{h_{peau}(C)}{h_{total}(C)} \quad (1)$$

A partir de ce rapport d'histogrammes calculé pour chaque pixel, nous pré-segmentons l'image en conservant les pixels pour lesquels  $p(peau|C)$  est supérieur à un seuil préalablement fixé (par exemple  $p(peau|C) > 0.5$ ). La figure 4 montre un exemple de pré-segmentation ainsi que la carte des probabilités associée.

Sur la première ligne, on note que quelques pixels proches de la couleur peau sont également sélectionnés. L'exemple de la deuxième ligne montre clairement une image pré-segmentée peu exploitable et qui justifie le rôle de la deuxième phase décrite ci-dessous.

### 3.3 Segmentation en régions

La segmentation en régions a pour but de déterminer les régions d'une image cohérentes à la fois spatialement et

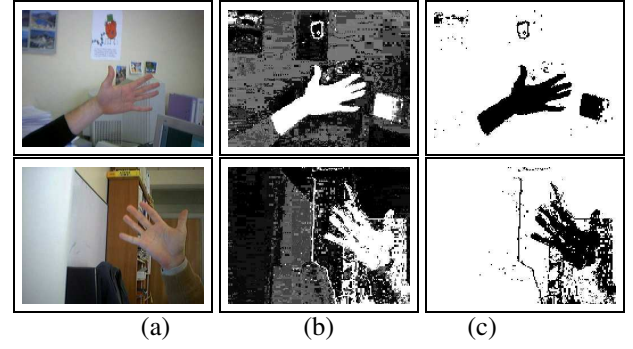


FIG. 4 – Exemple de pré-segmentation: (a)Image originale, (b)Carte des probabilités de couleur peau pour chaque pixel, (c) Image pré-segmentée

du point de vue de leur contenu. Dans notre approche, nous utilisons séquentiellement une clusterisation de l'espace de chrominance puis de luminance dans un algorithme basé sur le calcul de LPE. Après avoir utilisé une connaissance *a priori* de la couleur de la peau, nous utilisons ici les caractéristiques propres à l'image afin de séparer en différentes régions les pixels pré-segmentés.

### 3.4 Segmentation sur la chrominance

Notre approche débute par un *clustering* sur les composantes  $I_2$  et  $I_3$  à partir de l'histogramme des pixels peau préalablement détectés. Considérant cet histogramme bidimensionnel comme une image multi-niveaux de gris, nous appliquons un algorithme de LPE pour isoler les différents modes de couleurs présents dans l'image pré-segmentée des pixels peau.

Comme dans [1], nous utilisons un critère plus restrictif afin de sélectionner les marqueurs qui initialisent le calcul de LPE dans notre algorithme. Avant de rechercher ces marqueurs, un lissage préalable de l'histogramme est donc effectué afin de filtrer certains maxima locaux. Ce filtrage consiste en une simple dilation de l'histogramme par un élément structurant de taille  $4 \times 4$ . Cette taille correspond à la moyenne des écart-types calculés sur les composantes  $I_2$  et  $I_3$ , typiquement  $\sigma_{I_2} = 5.3$  et  $\sigma_{I_3} = 2.5$  dans nos expérimentations.

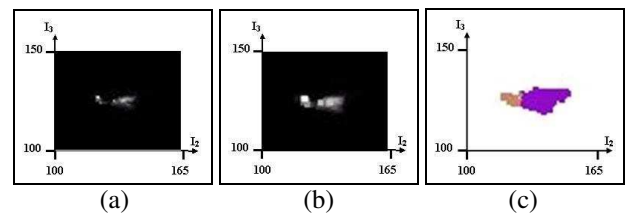


FIG. 5 – Histogrammes de la figure 4.(c haut) : (a)Histogramme de chrominance, (b)Histogramme dilaté, (c)Histogramme clusterisé

Comme indiqué précédemment, le choix des marqueurs pour le clustering à base de LPE est une étape très im-

portante. Bien que le pré-traitement de l'histogramme permette d'éliminer un nombre conséquent de maxima locaux parasites, nous calculons pour chaque maximum un critère appelé *contraste normalisé* qui permet de filtrer les pics parasites restants [1]. Ce critère permet d'éviter une sur-segmentation de l'image. Le contraste normalisé est défini pour chaque pic par :

$$\text{contraste\_normalise} = \frac{\text{Contraste}}{\text{Hauteur}} \quad (2)$$

où Hauteur désigne la hauteur du pic considéré et Contraste est égal à la différence de hauteur entre ce pic et la vallée qui le sépare du pic voisin le plus élevé. La sélection des maxima est réalisée par simple seuillage de la fonction *contraste\_normalise*, typiquement un maximum de contraste normalisé supérieur à 10% est conservé et constitue un marqueur pour l'algorithme de LPE. La figure 6 montre les Hauteurs et Contrastes pour différents pics d'un histogramme. L'axe parallèle à l'axe des niveaux de gris identifie les marqueurs retenus.

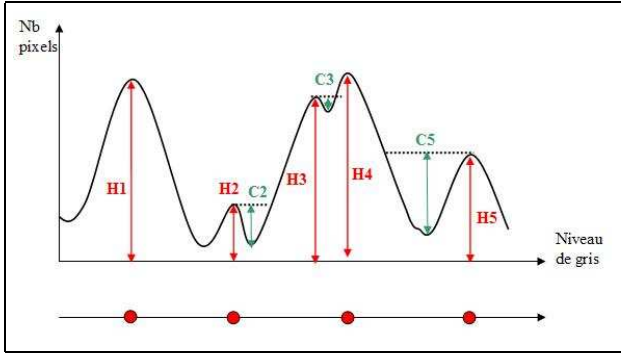


FIG. 6 – Hauteurs et Contrastes pour les pics d'un histogramme

L'exécution de l'algorithme de LPE à partir des marqueurs ainsi définis permet de clusteriser l'histogramme de chrominance. Les classes résultantes constituent les régions extraites. La figure 7 montre les régions ainsi obtenues sur les images vues précédemment. La figure de droite montre clairement un exemple où la région pertinente n'a pu être isolée de l'arrière-plan sur la base des composantes chromatiques. Nous nous proposons d'appliquer la même démarche sur les régions obtenues mais en exploitant la composante  $I_1$ .

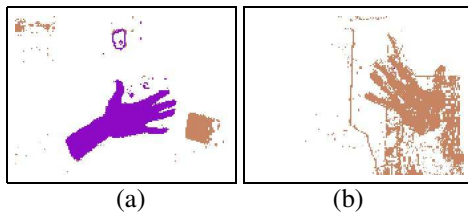


FIG. 7 – Exemple de clusterisation par chrominance

### 3.5 Segmentation par la luminance

La méthode reste la même que pour la chrominance mais nous l'appliquons aux régions précédemment caractérisées. Sur chacune de ces régions, nous établissons l'histogramme de l'intensité. La recherche des marqueurs s'effectue comme précédemment.

Ainsi, l'image segmentée par la chrominance est ici éventuellement sur-segmentée à l'aide de la luminance. La figure 8 présente l'effet de cette nouvelle segmentation sur l'exemple de la figure 7(b). Cette dernière passe n'a par contre aucun effet sur l'exemple 7(a), comme le montre la figure 8(a).

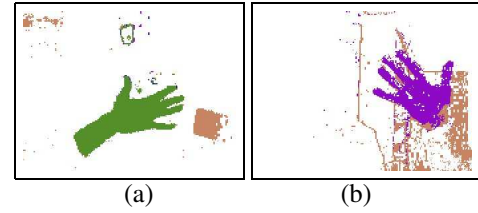


FIG. 8 – Exemple de clusterisation par chrominance et luminance

### 3.6 Elimination des petites régions

Le principe est de filtrer les régions inférieures à une taille donnée. Ce filtrage permet notamment d'éliminer des petites régions parasites connexes à la régions d'intérêt et isolées par les algorithmes successifs de LPE. Des résultats de segmentation incluant cette étape de filtrage sont montrés ci-après.

### 3.7 Exemples de segmentation

La figure 9 montre trois exemples de segmentation couleur. Nous constatons que les régions de couleur peau (main et/ou tête) sont relativement bien caractérisées même si certaines régions de l'arrière plan (souvent encombré) n'ont pu être filtrées. Dans l'exemple 9(a), la manche a pu être séparée de la main par le dernier LPE sur la luminance.

### 3.8 Exploitation de la segmentation

Le principe d'exploitation du résultat de cette segmentation dans le procesus d'estimation repose sur l'extraction d'un gradient couleur et la génération d'une image distance. Ces deux points sont repris ci-après.

**Gradient couleur.** Parmi les variantes de gradient couleur, nous avons préféré utiliser l'approche purement vectorielle proposée par Di Zenzo [20]. En effet, contrairement aux autres méthodes qui déterminent le gradient d'une multi-image comme fonction des vecteurs gradient sur chaque composante, cet opérateur extrait directement le gradient couleur et permet d'obtenir des contours moins nombreux qu'en utilisant un gradient classique. Rappelons brièvement le principe. Soient  $J$  la matrice Jacobienne et  $G = J'J$  la matrice carrée définies comme suit :

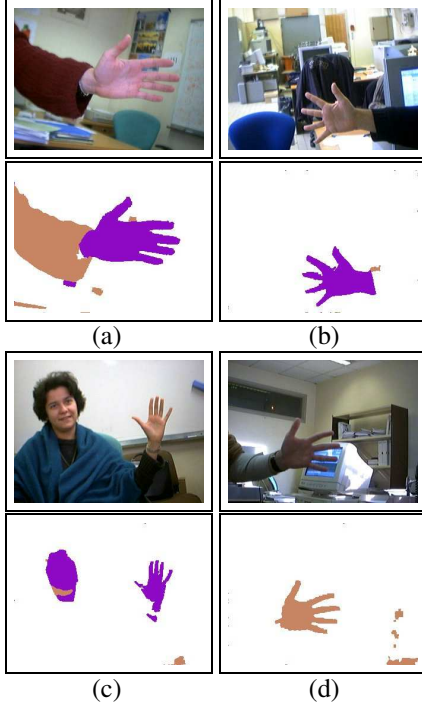


FIG. 9 – Exemples de segmentation couleur

$$J = \begin{pmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \\ \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} \\ \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{pmatrix}, G = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{pmatrix} = J^t J.$$

La direction du gradient de Di Zenzo est donnée classiquement par le vecteur propre qui correspond à la plus grande valeur propre de  $G$  tandis que son amplitude est la racine carrée de cette valeur propre. Avec ces notations, le calcul du gradient revient alors à trouver la valeur propre  $\lambda$  qui maximise l'équation

$$F(\theta) = g_{xx} \cdot \cos^2(\lambda) + 2 \cdot g_{xy} \cos(\lambda) \cdot \sin(\lambda) + g_{yy} \cdot \sin^2(\lambda),$$

soit, en posant  $\lambda' = 2 \cdot \theta$ ,

$$\lambda' = \arctan\left(\frac{2 \cdot g_{xy}}{g_{xx} - g_{yy}}\right), \lambda' \in \left]-\frac{1}{2}, \frac{1}{2}\right[.$$

La figure 10 montre un exemple de gradients extraits par Di Zenzo et, à titre de comparaison, de gradients extraits par Canny pour la même image.

**Génération de l'image de distance.** Le principe est ici de masquer les gradients couleur extraits de l'image par les contours de l'image des régions issues de la segmentation précédente afin de privilégier les contours pertinents (donc associés à la cible). Le résultat est une image de distance [4] construite à partir du gradient couleur mais dans laquelle les gradients masqués sont pénalisés. Ces gradients masqués ne sont donc pas complètement éliminés afin de pallier la segmentation parfois incomplète de la structure d'intérêt.

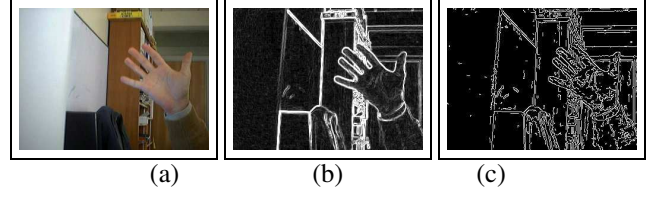


FIG. 10 – (a) Image originale, (b) Gradients extraits par Di Zenzo [20], (c) Gradients extraits par Canny

La génération puis l'exploitation dans le processus d'estimation (section 4.2) d'une image de distance permet de lisser l'image de contours et reste très appropriée pour évaluer à faible coût le critère pour chacune des  $N$  particules du filtre ( $N \gg 100$ ). La figure 11 présente l'image de distance ainsi générée pour l'image originale de la figure 10. Nous constatons que les gradients couleur correspondant aux contours des régions segmentées (figure 11(b)) sont bien privilégiés dans l'image de distance finale (faible intensité).

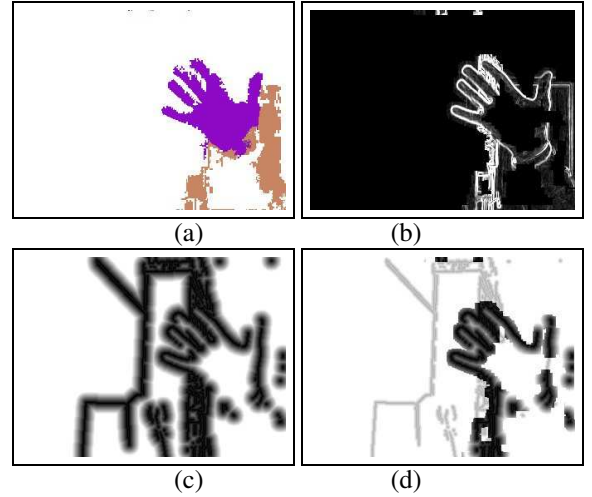


FIG. 11 – (a) Résultat de la segmentation couleur de la figure 10(a), (b) Contours de la figure 10(a) masqués par (a), (c) Image de distance, (d) Image de distance après masquage

## 4 Estimation par filtrage particulaire

### 4.1 Formalisme

Le filtre particulaire, ou algorithme de “Condensation” – pour “Conditional Density Propagation” –, est un estimateur de l'état d'un système non linéaire Markovien soumis à des excitations aléatoires possiblement non Gaussiennes. Le but est d'estimer récursivement la densité de probabilité a posteriori du vecteur d'état  $X_k$  à l'instant  $k$  conditionné sur la connaissance du processus de mesure jusqu'à cet instant, et de permettre le calcul de moments de cette distribu-

tion : moyenne conditionnelle de  $X_k$ , covariance conditionnelle, etc.

À chaque instant  $k$ , la densité  $p(X_k|z_1^k)$  est décrite par un ensemble de particules  $x_k^i$ , réalisations du vecteur d'état, affectées de poids  $w_k^i$ . Le problème est donc ramené à l'échantillonnage des valeurs  $x_k^i$  et à la définition des poids  $w_k^i$ , tout au long des étapes du processus d'estimation. Dès lors que ceci est acquis, la formule suivante, où  $N_i$  désigne le nombre de particules, permet l'approximation de moments de la distribution a posteriori de l'état :

$$E(g(X_k|z_1^k)) = \sum_{i=1}^{N_i} w_k^i g(x_k^i). \quad (3)$$

Cette approximation est d'autant meilleure que le nombre  $N_i$  est élevé.

L'initialisation de l'estimateur consiste en la création d'un ensemble de particules pondérées décrivant la connaissance *a priori*  $p(X_0)$ .

Disposant d'un ensemble de particules pondérées  $\{x_{k-1}^i, w_{k-1}^i\}$  associé à la densité de probabilité du vecteur d'état à l'instant  $k-1$  conditionné sur la connaissance des mesures jusqu'à cet instant, la détermination de la représentation de la densité  $p(X_k|z_1^{k-1})$  relative à la prédiction à l'instant suivant  $k$  procède en deux étapes.

La première étape consiste en le rééchantillonnage, dans l'ensemble  $\{x_{k-1}^i\}$ , de nouvelles particules  $x_{k-1}^{i'}$  en nombre demeurant égal à  $N_i$ . Ce rééchantillonnage est guidé par les valeurs des poids, au sens où une particule  $x_{k-1}^i$  admettant un poids  $w_{k-1}^i$  élevé peut être dupliquée plusieurs fois, au détriment d'une particule admettant un poids très faible qui est alors amenée à disparaître. L'ensemble  $\{x_{k-1}^{i'}, N_i^{-1}\}$  des particules uniformément pondérées résultant de cette redistribution demeure une approximation de  $p(X_{k-1}|z_1^{k-1})$ .

Ensuite, chaque particule  $x_{k-1}^{i'}$  est propagée entre les instants  $k-1$  et  $k$  par tirage aléatoire selon la distribution  $p(X_k|X_{k-1} = x_{k-1}^{i'})$  relative à la dynamique du vecteur d'état. Son poids est inchangé durant la propagation, de sorte que l'ensemble ainsi obtenu  $\{x_k^i, N_i^{-1}\}$  décrit effectivement la distribution  $p(X_k|z_1^{k-1})$ .

Enfin, la détermination de la représentation associée à  $p(X_k|z_1^k)$  est établie par une mise à jour de  $\{x_k^i, N_i^{-1}\}$  prenant en compte la réalisation  $z_k$  de la mesure.

D'après la loi de Bayes, on montre que la transformation de l'ensemble de particules pondérées  $\{x_k^i, N_i^{-1}\}$  en l'ensemble  $\{x_k^i, w_k^i\}$  permet de décrire la distribution a posteriori recherchée dès lors que l'on affecte à chaque poids  $w_k^i$  relatif à  $x_k^i$  la valeur  $p(z_k|X_k = x_k^i)$ , préalablement à une normalisation des  $w_k^i$  de façon que la somme  $\sum_{i=1}^{N_i} w_k^i$  soit égale à 1.

## 4.2 Application de la condensation au suivi d'une forme prototype

Comme indiqué précédemment, le suivi consiste à recaler une forme prototype sur l'image couleur acquise de la scène. Dans un premier temps, nous nous plaçons dans l'hypothèse où cette forme prototype est rigide et adaptée à la configuration de main que l'on veut suivre. Le but est donc d'estimer les paramètres globaux de la spline représentant cette forme, i.e. ses coordonnées de position, son facteur d'échelle ainsi que son orientation. L'ensemble de ces quantités est par conséquent regroupé dans le vecteur d'état

$$\vec{x} = [x, y, s, \theta]^t$$

On admet que la dynamique du vecteur d'état peut être décrite par le modèle auto-régressif suivant

$$\vec{X}_k = \mathbf{A}\vec{X}_{k-1} + \vec{W}_k, \quad (4)$$

$$\vec{X}_k = \begin{bmatrix} \vec{x}_k \\ \vec{x}_{k-1} \end{bmatrix}, \quad (5)$$

où l'indice  $k$  est relatif à la  $k^{\text{ème}}$  image de la séquence et  $W_k$  désigne le bruit de dynamique. Ce modèle est utilisé durant l'étape de prédiction de l'algorithme de condensation.

Dans le contexte considéré ici, le modèle de mesure doit permettre de pondérer une particule plus ou moins fortement selon que la spline qui lui est associée peut plus ou moins bien être mise en correspondance avec une forme d'intérêt dans l'image

Le traitement de l'image couleur brute a été conçu de telle sorte qu'il génère une image de distance dans laquelle les points de faible intensité sont certainement situés au voisinage des contours de la forme d'intérêt, les points de forte intensité en sont certainement éloignés, et les points d'intensités intermédiaires peuvent potentiellement être considérés comme points de contours.

Pour définir la densité de probabilité  $p(z_k|x_k^i)$  de l'observation à l'instant  $k$  conditionnée sur le fait que le vecteur d'état se réalise en la particule  $x_k^i$ , il suffit par exemple de répartir uniformément  $N$  points sur la spline paramétrisée par  $x_k^i$  et de poser

$$p(z_k|x_k^i) \propto \exp\left(-K \sum_{j=0}^N d^i(j)\right), \quad (6)$$

où  $d^i(j)$  désigne la valeur de l'intensité du pixel de l'image de distance correspondant au  $j^{\text{ème}}$  point,  $j = 1, \dots, N$ .

Du fait que l'on souhaite non seulement réaliser l'appariement d'une forme prototype, mais aussi sélectionner la forme la plus vraisemblable dans un ensemble défini a priori, l'algorithme utilisé précédemment pour le suivi est légèrement modifié de façon à intégrer une reconnaissance élémentaire du geste. Le vecteur d'état est augmenté par une composante discrète  $n$  relative à l'index de la configuration dans la batterie de configurations canoniques envisagées, soit :

$$\vec{x} = [n, x, y, s, \theta]^T, \text{ avec } n = 1, \dots, 4$$

Durant chaque étape de prédiction, cette composante est tirée aléatoirement en considérant que les configurations sont équiprobables. Le complément de l'étape de prédiction ainsi que l'étape de mise à jour sont inchangés. À chaque instant, le geste reconnu admet pour index celui dont le poids –calculé par sommation sur toutes les particules de même première composante– est le plus élevé.

## 5 Résultats expérimentaux

Afin d'évaluer les performances de l'algorithme, une mesure quantitative des taux de reconnaissance obtenus à partir de séquences de 800 images a été réalisée. Le développement a été effectué sous Linux à partir des bibliothèques OpenCV sur un PC équipé d'un processeur Pentium IV cadencé à 1.7GHz. Le filtre particulaire est composé de 500 particules. Dans cette configuration, l'algorithme s'exécute à une fréquence moyenne de 10Hz.

Notre algorithme a été évalué en filmant diverses personnes. Ces sujets sont donc différents du sujet qui a permis l'apprentissage des formes du *template*. La rigidité du template et la relative inadéquation de la forme des *templates* avec le contenu des images ne semble pas altérer les performances de notre système.

### 5.1 Exemple de suivi sur la main

La figure 12 montre le recalage et les commutations éventuelles du *template* sur quelques images intermédiaires d'une séquence de 400 images.

### 5.2 Exemple de suivi sur la silhouette

La figure 13 montre le recalage du *template* sur quelques images intermédiaires d'une séquence de 250 images.

Les tableaux 1 et 2 dressent un comparatif des taux de reconnaissance associés à chaque geste avec et sans l'utilisation du critère de couleur dans l'estimation.

TAB. 1 – Taux de reconnaissance avec arrière plan peu perturbé

|            | sans couleur | avec couleur |
|------------|--------------|--------------|
| config. #1 | 65 %         | 99.5 %       |
| config. #2 | 63 %         | 99 %         |
| config. #3 | 65 %         | 99.5 %       |
| config. #4 | 64 %         | 98 %         |

Le tableau 1 montre clairement que les taux de reconnaissance sont améliorés lorsque le résultat de la segmentation couleur est prise en compte dans la condensation. Le tableau 2 montre les taux de reconnaissance mesurés pour des séquences dont l'arrière plan est encombré.

Le tableau 3 donne des résultats moyens sur l'ensemble des configurations.

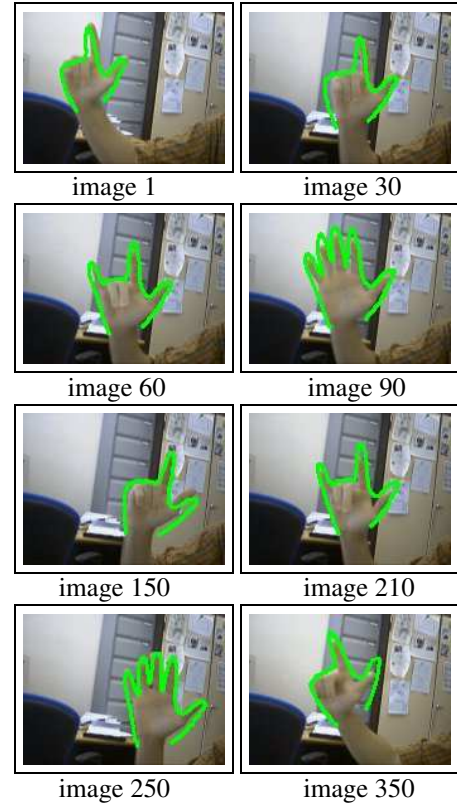


FIG. 12 – Exemples de recalage du *template* sur quelques images intermédiaires d'une séquence de 400 images

TAB. 2 – Taux de reconnaissance avec arrière plan perturbé

|            | sans couleur | avec couleur |
|------------|--------------|--------------|
| config. #1 | 45 %         | 99.5 %       |
| config. #2 | 43 %         | 93 %         |
| config. #3 | 44 %         | 98 %         |
| config. #4 | 40 %         | 89 %         |

TAB. 3 – Taux moyen de positionnement et de reconnaissance du geste

|                        | sans couleur | avec couleur |
|------------------------|--------------|--------------|
| pos correcte           | 86 %         | 99.5 %       |
| pos et config correcte | 50 %         | 95 %         |

## 6 Conclusion et perspectives

Nous avons présenté une méthode de suivi de gestes et de reconnaissance élémentaire de configurations de la main. Notre système se restreint à des mouvements qui sont grossièrement fronto-parallèles.

Dans un contexte de prises de vues depuis une caméra embarquée sur un robot mobile, les scènes observées sont a priori encombrées et sujettes à des conditions de luminosité variables.

Aussi, et pour s'affranchir au mieux des fausses mesures, la méthode proposée combine des informations de forme et de couleur dans une méthode de segmentation origi-

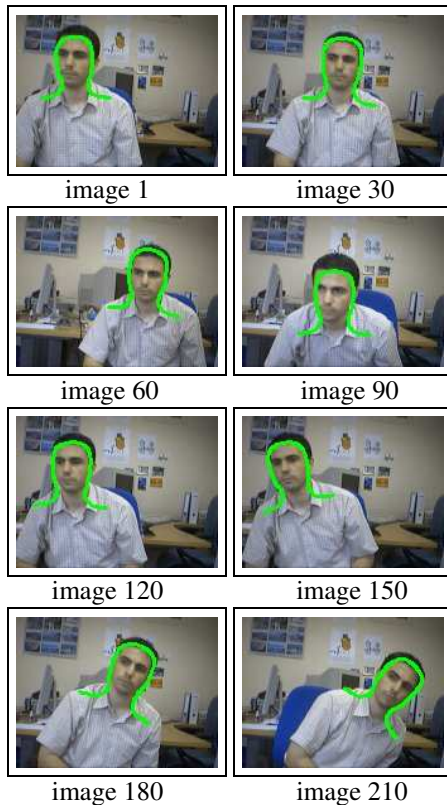


FIG. 13 – Exemples de recalage du *template* sur quelques images intermédiaires d’une séquence de 250 images

nale. Cette méthode consiste à détecter les pixels de la peau puis à segmenter en régions à partir d’un algorithme de LPE sur la chrominance et la luminance des pixels ainsi sélectionnés. Le masquage des gradients couleur de l’image par le résultat de cette segmentation en régions permet de privilégier les contours les plus pertinents.

Le recalage du *template* et la sélection du *template* le plus vraisemblable s’effectuent, à partir de ces contours, dans un seul et unique processus de condensation, permettant ainsi d’espérer des temps de calcul compatibles avec les applications visées.

Les premiers résultats obtenus sont encourageants : le suivi est assez robuste aux fausses mesures générées notamment par l’arrière plan de la scène tandis que les commutations d’un *template* à un autre sont relativement bien détectées. Un filtrage temporel devrait éliminer les commutations intempestives.

Cette nouvelle fonctionnalité pour l’interaction sera prochainement intégrée sur nos plateformes mobiles. Elle permettra en outre aux usagers qui partagent l’environnement du robot de le commander gestuellement afin de faciliter/sécuriser ses déplacements en présence de public.

À terme, et pour enrichir cette interaction, des aspects dynamiques seront considérés dans la reconnaissance. Le système devra, comme ébauché dans cet article, gérer plusieurs hypothèses quant à la forme du *template* mais

également la nature de sa trajectoire image (modèles canoniques de déplacements, connaissance a priori sur les possibilités d’enchaînements, ...).

## Références

- [1] A. Albiol, L. Torres, and Ed. Delp. An unsupervised color image segmentation algorithm for face detection applications. *IEEE International Conference on Image Processing (ICIP'01)*, pages 7–10, October 2001.
- [2] K.H. Quek B.D. Zarit, B.J. Super. Comparison of five color models in skin pixel classification. In *ICCV'99 International Workshop on Recognition, Analysing, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'99)*, pages 58–63, Corfu, Greece, September 1999.
- [3] S. Beucher and F. Meyer. *Mathematical Morphology in Image Processing, Chapter 12*. Marcel Dekker Inc., 1993.
- [4] G. Borgefors. Distance Transforms in Digital Images. *IEEE Trans. Pattern Analysis Machine Intelligence (PAMI'86)*, 8:344–371, 1986.
- [5] L. Bretzner, I. Laptev, and T. Lindeberg. Hand Gesture using Multi-scale Colour Features, Hierarchical Models and Particle Filtering. In *Int. Conf. on Automatic Face and Gesture Recognition (FGR'02)*, pages 405–410, Washington, 2002.
- [6] L. Davis, V. Philomin, and R. Duraiswani. Tracking Humans from a Moving Platform. In *Int. Conf. on Pattern Recognition (ICPR'00)*, volume 4, pages 171–177, Barcelona, 2000.
- [7] Q. Delamarre and O. Faugeras. 3D Articulated Models and Multi-view Tracking with Physical Forces. Technical report, Institut National de Recherche en Informatique et Automatique (INRIA), Sophia-Antipolis, 2001.
- [8] D.M. Gavrilă. The visual analysis of human movement : A survey. *Computer Vision and Image Understanding (CVIU'99)*, (1):82–98, January 1999.
- [9] D.M. Gavrilă. Pedestrian Detection from a Moving Vehicle. In *European Conf. on Computer Vision (ECCV'00)*, Dublin, June 2000.
- [10] M. Isard and A. Blake. Conditional Density Propagation for Visual Tracking. *Int. Journal of Computer Vision (IJCV'98)*, 29(1):5–28, 1998.
- [11] I. Kakadiaris, D. Metaxas, and R. Bajcsy. Active Part Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-based Approach. In *Computer Vision and Pattern Recognition (CVPR'94)*, pages 35–46, Seattle, 1994.
- [12] C. Kervrann and F. Heitz. Apprentissage non supervisé et suivi de modèles déformables dans une séquence d’images. In *Reconnaissance des Formes*

*et Intelligence Artificielle (RFIA'96)*, Rennes, Janvier 1996.

- [13] Y. Ohta, T. Kanade, and T. Sakai. Color Information for Region Segmentation. *Computer Graphics and Image Processin (CGIP'80)*, (13):222–241, 1980.
- [14] N. Paragios and R. Deriche. Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects. *IEEE Trans. Pattern Analysis Machine Intelligence(PAMI'00)*, 22(3):850–863, 2000.
- [15] V. Pavlovic, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI'97)*, 19(7):677–695, 1997.
- [16] J.M. Rehg and T. Kanade. Visual Tracking of High DOF Articulated Structures : Application to Human Hand Tracking. In *European Conf. on Computer Vision (ECCV'96)*, pages 35–46, 1996.
- [17] K. Schwerdt and J.L. Crowley. Robust Face Tracking using Color. In *Int. Conf. on Face and Gesture Recognition (FGR'00)*, pages 90–95, Grenoble, March 2000.
- [18] B. Stenger, P. R. S. Mendonça, and R. Cipolla. Model based 3D Tracking of an Articulated Hand. In *Computer Vision and Pattern Recognition (CVPR'01)*, volume 2, December 2001.
- [19] D. Terzopoulos and D. Metaxas. Dynamic 3D Models With Local and Global Deformations : Deformable Superquadrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI'91)*, 13(7):703–714, July 1991.
- [20] S. Di Zenzo. A Note on the Gradient of a Multi-Image. *Int. Journal of Computer Graphics and Image Processing (CVGIP'86)*, 33:116–125, 1986.