

Segmentation automatique de personnes par coupe de graphe et gabarits

Cyrille MIGNIOT, Pascal BERTOLINO, Jean-Marc CHASSERY

GIPSA-lab - Grenoble INP - France
961 rue de la Houille Blanche , BP 46 ,38402 Grenoble Cedex, France
prenom.nom@gipsa-lab.grenoble-inp.fr

Résumé – Ce papier présente une nouvelle méthode de segmentation automatique de personnes dans les images. Elle se base sur la méthode efficace de la coupe de graphe en l’adaptant pour mieux correspondre à la classe des personnes. En effet, ni la couleur ni la texture ne sont discriminantes pour cette classe et ne peuvent donc être utilisées. En revanche, la forme de la silhouette est pertinente. La coupe de graphe est alors guidée par un gabarit non binaire représentant la probabilité de chaque pixel de l’image d’appartenir à la personne à segmenter. Dans un second temps, un nouveau gabarit réalisé par parties est calculé pour envisager différentes postures possibles de la personne sans trop augmenter la taille des calculs et ainsi rester proche du temps réel.

Abstract – This paper presents a new fully automatic method for segmenting upright people in the images. It is based on the efficient graph cut segmentation. Since colour and texture prevent from discriminating this particular class, silhouette shape is used instead. The graph cut is guided by a non-binary template of silhouette that represents the probability of each pixel to be a part of the person to segment. Subsequently, part-based template is used to better take into account the different postures of a person. Our method is close to real time and is tested on a large person dataset.

1 Introduction

Pour la détection et la segmentation de personnes, de nombreuses méthodes utilisent des gabarits binaires. Un gabarit est un modèle permettant de caractériser la forme générale des éléments d’une classe. Dans celle des personnes, les postures pouvant apparaître sont assez variées et influencent vraiment la segmentation. Un gabarit représente alors l’allure de la silhouette pour une de ces postures par un masque binaire. Un catalogue de gabarits est réalisé contenant toutes les postures que la personne peut prendre. Ces gabarits sont ensuite comparés un par un aux caractéristiques de l’image (souvent les contours [14]). Si la comparaison est positive, une personne est détectée et le gabarit donne sa posture. La segmentation est finalement obtenue en adaptant légèrement la silhouette de ce gabarit aux contours de l’image [10]. Pour diminuer le temps de calcul, Gavrilin et al [5] réalisent une répartition hiérarchique des gabarits. Enfin, Lin et al [6] décomposent le corps en trois parties (le torse, le bassin et les jambes) et cherchent le sous-gabarit correspondant à chacune de ces parties.

La coupe de graphe, telle que définie par Boykov et al [1], est une méthode efficace de segmentation. Assez simple, elle possède également l’avantage de permettre une interaction facile avec l’utilisateur. Elle est donc couramment utilisée en segmentation d’images [11]. Le graphe considéré est constitué d’une source et d’un puits correspondant au premier et à l’arrière plan et de noeuds correspondant aux pixels de l’image. Des arêtes de voisinage relient les pixels voisins spatialement

et des arêtes de liaisons relient les pixels au puits et à la source. Des pondérations sont associées aux arêtes. La coupe du graphe qui minimise la somme des pondérations des arêtes coupées (ou flot) est alors calculée et sépare les pixels du premier et de l’arrière plan. Les pondérations des arêtes de voisinage sont reliées aux contours de l’image alors que les pondérations des arêtes de liaison sont reliées à la probabilité du pixel d’appartenir au premier ou à l’arrière plan (généralement sur un critère de couleur, grâce aux indications de l’utilisateur).

Un certain nombre de travaux ont déjà été réalisés pour introduire des contraintes de forme dans la coupe de graphe. Un troisième terme a été rajouté par Freedman et al [4] à la fonction d’énergie pour prendre en compte la forme à partir de lignes de niveau. Le terme de région ou de contour (ou les deux) ont été modifiés dans plusieurs travaux [13, 12, 3]. Malcolm et al [7] ont proposé une solution intéressante mais coûteuse en temps de traitement où une pré-image obtenue par apprentissage avec une ACP à noyau est itérativement réactualisée.

Nous introduisons une nouvelle technique pour adapter la coupe de graphe aux caractéristiques des personnes sans interaction de l’utilisateur. Seule la segmentation est prise en compte, la détection préliminaire des personnes étant effectuée par la méthode de Dalal et al [2] qui fournit des fenêtres normalisées centrées sur la personne (figure 5). Notre contribution est dans un premier temps de pondérer le graphe par la silhouette moyenne obtenue sur une base d’apprentissage et appelée gabarit, et ensuite d’adapter cette technique avec un gabarit

par parties construit par des coupes de graphe successives appliquées sur chaque partie du corps de la personne.

En section 2, nous introduisons un gabarit non binaire à partir d'une base de données d'apprentissage, qui représente la probabilité d'un pixel d'appartenir à la silhouette de la personne. Cette probabilité est utilisée pour initialiser la pondération des arêtes de liaison.

Ensuite, en section 3, la segmentation est affinée pour s'adapter aux différentes postures : l'image et donc le corps sont divisés en plusieurs parties. Pour chaque partie, plusieurs sous-gabarits sont testés et le gabarit final (appelé gabarit par parties) est obtenu par concaténation des meilleurs sous-gabarits.

Finalement, en section 4, nos deux approches (avec un gabarit unique ou un gabarit par parties) sont évaluées.

2 Création du graphe

Le graphe est réalisé à partir d'une fenêtre englobant une personne et issue de la détection. Une source F et un puits B représentent le premier et l'arrière plan. Chaque pixel est relié dans le graphe aux pixels voisins par des arêtes de voisinage et à F et B par des arêtes de liaisons. L'importance relative des pondérations des arêtes de voisinage et de liaison est réglée par deux coefficients (α et β).

2.1 Arêtes de voisinage

Les arêtes de voisinage représentent la possibilité que la transition entre deux pixels voisins soit sur la circonférence de la silhouette découpée par la coupe de graphe. Elles sont donc logiquement associées aux contours de l'image. La différence d'intensité est utilisée comme dans la méthode de Boykov [1]. Soit I_p l'intensité du pixel p et I_q l'intensité du pixel q . La pondération associée à l'arête entre p et q est définie par :

$$\omega_{pq} = \beta e^{-\frac{|I_p - I_q|^2}{2\sigma^2}} \quad (1)$$

où σ réalise le filtrage opéré par l'exponentielle. Les valeurs faibles représentent les fortes probabilités de contour.

2.2 Arêtes de liaison

Les arêtes de liaison rattachent tous les pixels à F et B . Pour un pixel, la coupe passe par une et une seule de ces arêtes, ce qui désigne à quelle région est assigné le pixel. Leur pondération doit donc correspondre à la probabilité du pixel d'appartenir au premier ou à l'arrière plan. Boykov et al [1] relie cette probabilité à la distribution des couleurs. Mais la grande variété des couleurs dans la classe des personnes rend cette caractéristique peu discriminante et nécessiterait l'intervention d'un utilisateur. La forme de la silhouette est une information plus pertinente. On réalise alors un gabarit qui représente la probabilité t_p de chaque pixel p d'appartenir à une silhouette humaine. Ce gabarit est la moyenne d'un ensemble de 200 silhouettes représentatives des différentes postures debout (figure

1). Les pondérations attribuées aux arêtes de liaison sont alors définies par :

$$\omega_p^F = -\alpha \ln(t_p) \quad (2)$$

$$\omega_p^B = -\alpha \ln(1 - t_p) \quad (3)$$



FIG. 1: Image moyenne d'une base de 200 silhouettes humaines. Notons que la position de la tête et du torse est bien plus stable que celle des bras et surtout que celle des jambes.

3 Gabarit par parties

Le gabarit présenté précédemment prend en compte toutes les postures mais défavorise cependant celles d'occurrences les plus faibles (notamment lorsque les jambes sont écartées). L'idée est alors de choisir un gabarit qui corresponde à la posture rencontrée. Il serait possible de réaliser une coupe de graphe à partir de gabarits représentant toutes les postures possibles. Mais le nombre de postures étant élevé, on obtiendrait un temps de traitement très important. Nous proposons de partager l'image en cinq parties : la tête, les parties droite et gauche du torse incluant le bras, la jambe droite et la jambe gauche. Pour chacune de ces parties, un certain nombre de sous-gabarits sont construits pour les différentes postures possibles (figure 2). Pour chaque partie, une coupe de graphe est réalisée à partir de chacun des sous-gabarits. Le gabarit par

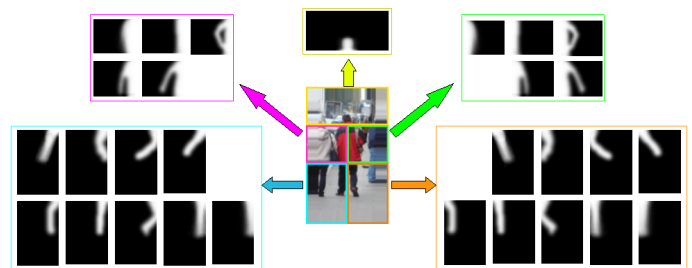


FIG. 2: L'image est découpée en cinq parties. Des sous-gabarits représentent les postures possibles dans chacune d'entre elles.

parties est la concaténation des sous-gabarits ayant donné le flot de coupe le plus faible pour chaque partie (figure 3). Enfin, une coupe de graphe sur l'image entière à partir de ce gabarit par parties permet d'assurer la continuité de la silhouette segmentée.

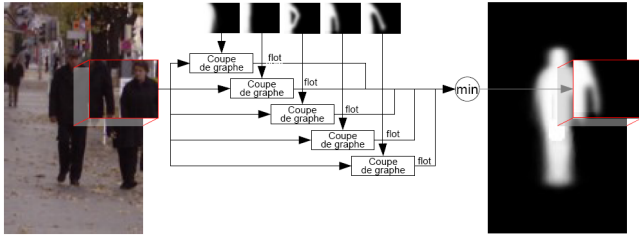


FIG. 3: Pour chaque partie de l'image, des coupes de graphe sont réalisées à partir de plusieurs sous-gabarits. Celui associé à la coupe donnant le flot minimal est alors ajouté au gabarit par parties.

Un nombre de sous-gabarits restreint est suffisant au traitement. En effet, comme les arêtes de voisinage adaptent la coupe aux contours, les gabarits doivent seulement favoriser un état (bras décollé, jambe pliée, ...) et non parfaitement correspondre à la silhouette de la personne.

4 Performances

Pour évaluer notre méthode nous avons réalisé des tests sur un ensemble de 400 images de personnes issues de la base de données statique de l'INRIA. La $F_{measure}$ et la mesure de Yasnoff [9], par comparaison avec une réalité terrain réalisée manuellement et disponible dans [8], nous permettent d'évaluer objectivement une segmentation. Les moyennes de ces deux mesures pour les segmentations de l'ensemble des images testées donnent alors une quantification des performances de notre méthode. Les temps de traitement sont observés à partir d'une implémentation C++ non optimisée sur un processeur Pentium D 3GHz.

4.1 Réglage optimal du procédé

Pour régler de façon optimale le procédé, nous avons déterminé les valeurs des paramètres α , β et σ qui produisent les meilleures segmentations. Les résultats affichés dans la figure 4 sont obtenus avec un gabarit par parties mais ceux obtenus avec un gabarit unique sont très semblables. Obtenir les meilleures segmentations revient à minimiser la mesure de Yasnoff et maximiser la $F_{measure}$. Nous choisissons de garder pour la suite les valeurs : $\sigma = 9$, $\alpha = 12$ et $\beta = 60$.

4.2 Gabarit unique ou par parties

Puisque nous avons introduit deux procédés différents, il faut comparer leurs performances respectives. Tout d'abord le traitement avec un gabarit unique est logiquement plus rapide avec (pour le traitement d'images 96×160 pixels) une moyenne de 12 ms par image contre une moyenne de 70 ms avec le gabarit par parties. Concernant la qualité de la segmentation, en utilisant les valeurs de paramètres fixées précédemment, on obtient les résultats du tableau 1 : le gabarit par parties donne

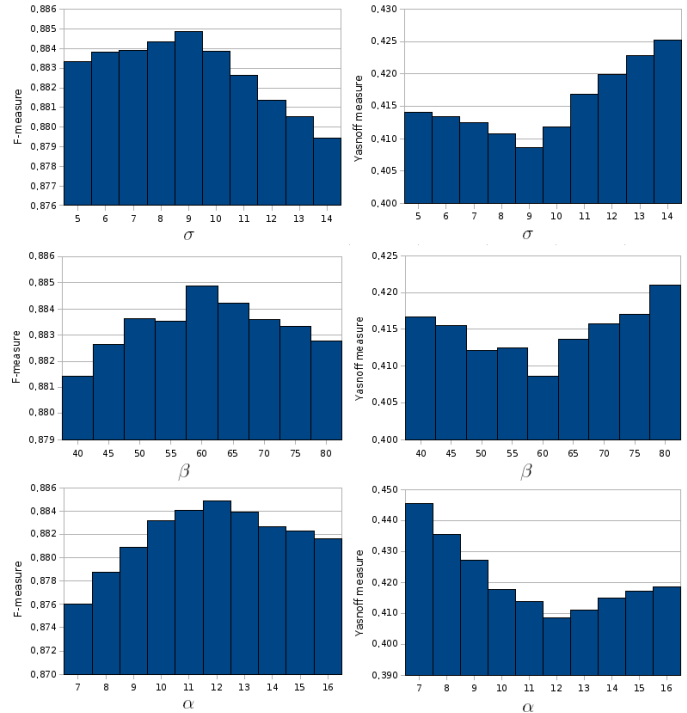


FIG. 4: Résultats des tests pour optimiser σ (première ligne), β (seconde ligne) et α (troisième ligne). La $F_{measure}$ (première colonne) et la mesure de Yasnoff (seconde colonne) donnent une évaluation de la qualité de la segmentation.

en moyenne une meilleure $F_{measure}$ et une meilleure mesure de Yasnoff. Dans la grande majorité des cas, le gabarit par parties réalise une segmentation visuellement mieux adaptée et plus précise (figure 5).

Mesure	Gabarit unique	Gabarit par parties
$F_{measure}$	0,8813	0,8849
Yasnoff	0,4178	0,4087

TAB. 1: Sur un ensemble de 400 tests et pour deux mesures objectives, le gabarit par parties donne de meilleurs résultats que le gabarit unique

5 Conclusion

Nous avons proposé une nouvelle méthode de segmentation adaptant la coupe de graphe traditionnelle au cas particulier des personnes. Pour cela, nous avons introduit des gabarits non binaires pour évaluer la localisation de la silhouette. Le gabarit est soit général à toute la classe soit adapté à la posture de la personne. Le traitement est efficace et proche du temps réel. L'étude de séquences vidéo ainsi que la possibilité d'une interaction facile avec l'utilisateur sont parmi les principaux avantages de la coupe de graphe. Des travaux futurs pertinents seraient alors d'adapter notre méthode aux vidéos et de permettre une interaction performante avec la classe des personnes.

References

- [1] Y.Y. Boykov et M.P. Jolly, Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images, *International Conference on Computer Vision*, 1:105-112, Juillet 2001.
- [2] N. Dalal et B. Triggs, Histograms of Oriented Gradients for Human Detection, *International Conference on Computer Vision & Pattern Recognition*, 2:886-893, Juin 2005
- [3] P. Das, O. Veksler, V. Zavadsky et Y. Boykov, Semiautomatic Segmentation with Compact Shape Prior, *Image and Vision Computing*, 27:206-219, Janvier 2009
- [4] D. Freedman et T. Zhang, Interactive Graph Cut Based Segmentation with Shape Priors, *International Conference on Computer Vision & Pattern Recognition*, 1:755-762, 2005
- [5] D.M. Gavrila et J. Giebel, Shape-Based Pedestrian Detection and Tracking, *Intelligent Vehicle Symposium*, 1:8-14, Juin 2002
- [6] Z. Lin, L.S. Davis, D. Doermann et D. DeMenthon, Hierarchical Part-Template Matching for Human Detection and Segmentation, *International Conference on Computer Vision*, 1-8, Octobre 2007
- [7] J. Malcolm, Y. Rathi et A. Tannenbaum, Graph Cut Segmentation with Nonlinear Shape Priors, *International Conference on Image Processing*, 4:365-368, Octobre 2007
- [8] C. Migniot, P. Bertolino et J.M. Chassery, Contour Segment Analysis for Human Silhouette Presegmentation, *International Conference on Computer Vision Theory and Applications*, Mai 2010
- [9] S. Philipp-Foliguet et L. Guigues, Mutlti-Scale Criteria for the Evaluation of Image Segmentation Algorithms, *Journal of Multimedia*, 3:42-56, 2008
- [10] M.D. Rodriguez et M Shah, Detecting and Segmenting Humans in Crowded Scenes, *International Conference on Multimedia*, 353-356, 2007
- [11] C. Rother, V. Kolmogorov et A. Blake, GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts, *ACM Transactions on Graphics*, 23:309-314, Août 2004
- [12] G.G. Slabaugh et G. Unal, Graph Cuts Segmentation using an Elliptical Shape Prior, *International Conference on Image Processing*, 2:1222-1225, Septembre 2005
- [13] H. Wang et H. Zhang, Adaptive Shape Prior in Graph Cut Segmentation, *International Conference on Image Processing*, 3029-3032, Septembre 2010
- [14] L. Zhao et L.S. Davis, Closely Coupled Object Detection and Segmentation, *International Conference on Computer Vision*, 1:454-461, Octobre 2005



FIG. 5: De gauche à droite : image initiale, segmentation obtenue avec un gabarit unique puis avec un gabarit par parties. Dans la plupart des cas, la segmentation est plus précise avec le gabarit par parties.