



HAL
open science

Segmentation automatique de personnes par coupe de graphe et gabarits

Cyrille Migniot, Pascal Bertolino, Jean-Marc Chassery

► **To cite this version:**

Cyrille Migniot, Pascal Bertolino, Jean-Marc Chassery. Segmentation automatique de personnes par coupe de graphe et gabarits. XXIIIe colloque GRETSI (GRETSI 2011), Sep 2011, Bordeaux, France. pp.CD. hal-00593222v1

HAL Id: hal-00593222

<https://hal.science/hal-00593222v1>

Submitted on 13 May 2011 (v1), last revised 17 Jun 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Segmentation automatique de personnes par coupe de graphe et gabarits

Cyrille MIGNIOT, Pascal BERTOLINO, Jean-Marc CHASSERY

Gipsa-lab UMR 5216 CNRS - Grenoble Universités - France
prenom.nom@gipsa-lab.grenoble-inp.fr

Thème – 4.2 - Segmentation et détection de ruptures

Problème traité – Segmentation automatique de personnes pré-détectées dans les images fixes

Originalité – La coupe de graphe est guidée par un ou plusieurs gabarits de silhouette

Résultats – Evaluation avec des métriques reconnues sur une base de 400 images. Des résultats réellement utilisables dans diverses applications

1 Contexte

La détection de personnes dans les images est une fonctionnalité de plus en plus répandue dans de nombreux domaines. La segmentation automatique de ces personnes est aussi un élément capital dans de nombreuses applications parmi lesquelles l'édition d'image, la vidéo-surveillance et la robotique. Obtenir une segmentation de qualité dans des environnements non contrôlés reste un problème ouvert. Pour la détection préalable de ces personnes, nous utilisons la méthode de Dalal et al [1] qui fournit des fenêtres normalisées centrées sur la personne. Pour détecter et segmenter ces personnes, de nombreuses méthodes utilisent un ou plusieurs gabarits binaires, comparés un par un aux caractéristiques de l'image [2]. Si la comparaison est positive, une personne est détectée et la silhouette correspondant à sa posture est évaluée. La segmentation est finalement obtenue en adaptant légèrement cette silhouette aux contours de l'image [3, 4]. La silhouette peut être décomposée en plusieurs parties auxquelles on cherche le gabarit le mieux adapté [5]. Dans tous les cas, les traitements sont lourds et il n'existe pas vraiment d'évaluation quantitative des résultats. Nous proposons une technique rapide que nous évaluons avec des métriques existantes. Notre technique est fondée sur la coupe de graphe [6, 7] qui est une méthode de segmentation efficace et que nous rappelons ici : le graphe est constitué de noeuds correspondant aux pixels de l'image. Des **arêtes de voisinage** relient ces noeuds pour modéliser l'adjacence des pixels. Deux noeuds particuliers appelés source et puits sont rajoutés au graphe afin de représenter le premier et l'arrière plan. Des **arêtes de liaison** relient chaque pixel au puits et à la source. Des pondérations sont associées à toutes les arêtes. La coupe du graphe qui minimise la somme des pondérations des arêtes coupées est alors calculée et sépare les pixels de premier et d'arrière plan. Afin de prendre en compte le contenu de l'image, les pondérations des arêtes de voisinage sont reliées au gradient 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 mais la plupart du temps en modifiant ou en rajoutant un terme à la fonction d'énergie à minimiser et en réalisant des traitements itératifs coûteux [8, 9, 10, 11, 12].

2 Initialisation du graphe avec un gabarit unique

Le graphe est réalisé à partir d'une fenêtre englobant une personne, issue de la détection. Une source F et un puits B représentent le premier et l'arrière plan. Chaque pixel de la fenêtre est relié aux pixels voisins par des arêtes de voisinage et à F et B par des arêtes de liaison. L'importance relative des pondérations des arêtes de voisinage et de liaison est réglée par deux coefficients (α et β). Les arêtes de voisinage représentent la possibilité que la transition entre deux pixels voisins soit un contour de la silhouette découpée par la coupe de graphe. Elles sont donc logiquement associées au gradient de l'image. Boykov et al [6] utilisent la différence d'intensité pour pondérer ces arêtes. Compte tenu de la petite taille des images étudiées et des discontinuités rapides

qu'on y trouve, il est préférable d'utiliser le module du gradient au pixel p noté g_p . La pondération associée à l'arête reliant deux pixels p et q est alors définie par :

$$\omega_{pq} = \alpha e^{-\frac{g_p^2}{2\sigma^2}} \quad (1)$$

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

Les arêtes de liaison rattachent chacun des pixels à F et B . Cette probabilité peut être reliée à la distribution des couleurs [6]. Mais la grande variété de couleurs dans la classe des personnes et dans le fond rend cette caractéristique peu discriminante et nécessiterait l'intervention d'un utilisateur. La forme d'une silhouette humaine 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 silhouette moyenne obtenue sur une base de 200 silhouettes représentatives des différentes postures debout (figure 1.a). Les pondérations attribuées aux arêtes de liaison sont alors définies par :

$$\omega_p^F = -\beta \cdot \ln(1 - t_p) \quad \omega_p^B = -\beta \cdot \ln(t_p) \quad (2)$$

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 d'utiliser un gabarit adapté à la posture rencontrée. Il serait possible de réaliser une coupe de graphe avec chaque gabarit représentant une des postures possibles. Mais le nombre de postures étant élevé, on obtiendrait un temps de traitement très important. Nous proposons de traiter indépendamment les différentes parties de la silhouette et de découper l'image en cinq parties : tête, partie droite du torse, partie gauche du torse, jambe droite et jambe gauche. Pour chacune de ces parties, un certain nombre de sous-gabarits sont construits à l'aide des postures correspondantes dans la base de données des personnes (figure 1.b). Pour chaque partie, une coupe de graphe est réalisée à partir de chacun des sous-gabarits. Le gabarit par parties est la concaténation des sous-gabarits ayant donné le flot de coupe le plus faible pour chaque partie. Enfin, une coupe de graphe sur l'image entière à partir du gabarit par parties obtenu donne la segmentation finale.

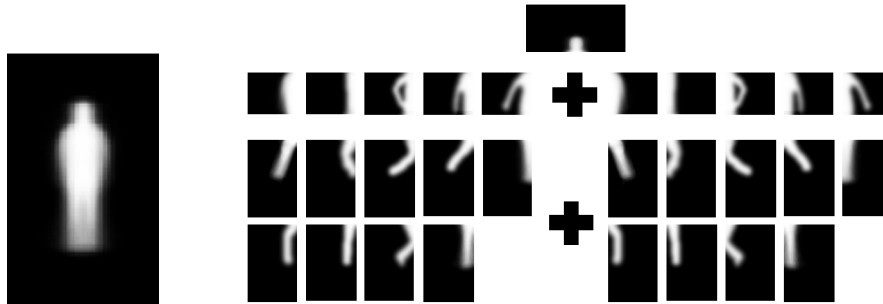


FIGURE 1 – (a) gabarit unique (b) Ensemble des sous-gabarits relatifs aux cinq parties de l'image

4 Evaluation

Le traitement d'une image 96×160 pixels prend une moyenne de 12 ms avec le gabarit unique contre 70 ms avec le gabarit par parties. 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 [13], par comparaison avec une réalité terrain réalisée manuellement et disponible dans [14], nous permettent à la fois de trouver les valeurs optimales pour α , β et σ et d'évaluer objectivement une segmentation. Le gabarit unique donne une meilleure mesure de Yasnoff alors que le gabarit par parties donne une meilleure $F_{measure}$. En analysant les différentes segmentations obtenues, on voit que parfois un sous-gabarit inadapté est utilisé à tort ce qui provoque des segmentations aberrantes auxquelles la mesure de Yasnoff est très sensible. C'est la raison pour laquelle la mesure de Yasnoff est meilleure avec un gabarit unique. Néanmoins, dans la grande majorité des cas, le gabarit par parties donne une segmentation mieux adaptée et plus précise (figure 2) ce qui explique une meilleure $F_{measure}$. En effet, si on supprime de la base de tests les images qui ont donné avec un gabarit par parties des valeurs de mesure de Yasnoff moins bonnes qu'avec un gabarit unique (ce que nous évaluons à seulement 2,5% puis 5% des images), on remarque (figure 3) que le gabarit par parties donne alors des résultats largement meilleurs. Dans la majorité des cas, le gabarit par parties est donc préférable.

Dans des travaux futurs, nous adapterons notre méthode aux vidéos avec la possibilité d'une interaction minimaliste avec l'utilisateur lorsque nécessaire.

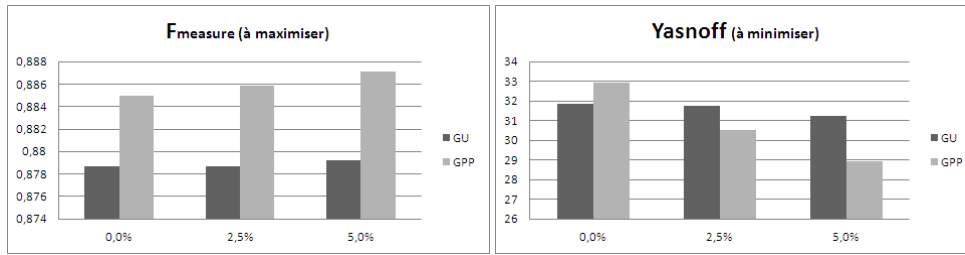


FIGURE 2 – $F_{measure}$ et mesure de Yasnoff avec un gabarit unique (GU) ou un gabarit par parties (GPP) en retirant 0%, 2, 5% ou 5% des images de la base de données (celles donnant les résultats aberrants)

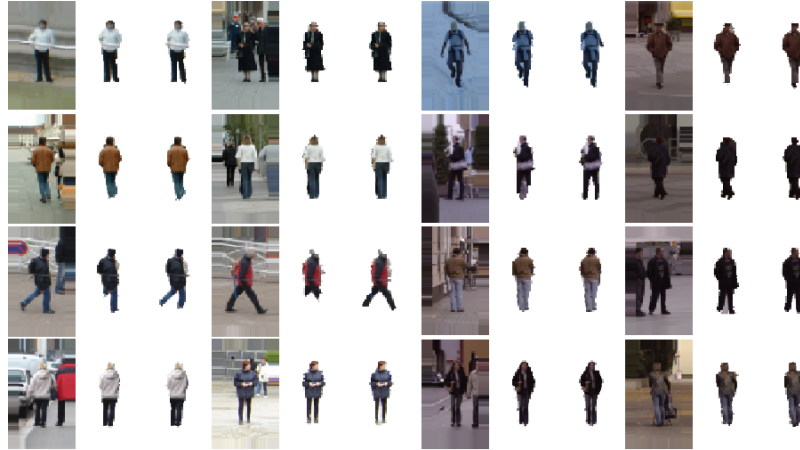


FIGURE 3 – De gauche à droite : image initiale, segmentation obtenue avec un gabarit unique et avec un gabarit par parties

Références

- [1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *CVPR*, vol. 2, pp. 886–893, June 2005.
- [2] L. Zhao and L.S. Davis, “Closely coupled object detection and segmentation,” in *ICCV*, 2005, vol. 1, pp. 454–461.
- [3] M.D. Rodriguez and M. Shah, “Detecting and segmenting humans in crowded scenes,” in *ICM*, 2007, pp. 353–356.
- [4] D.M. Gavrilu and J. Giebel, “Shape-based pedestrian detection and tracking,” *Intelligent Vehicle Symposium*, vol. 1, pp. 8–14, June 2002.
- [5] Z. Lin, L.S. Davis, D. Doermann, and D. DeMenthon, “Hierarchical part-template matching for human detection and segmentation,” *ICCV*, pp. 1–8, October 2007.
- [6] Y. Boykov and M.P. Jolly, “Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images,” *ICCV*, vol. 1, pp. 105–112, July 2001.
- [7] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut : Interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, pp. 309–314, August 2004.
- [8] D. Freedman and T. Zhang, “Interactive graph cut based segmentation with shape priors,” in *CVPR (1)*, 2005, pp. 755–762.
- [9] H. Wang and H. Zhang, “adaptive shape prior in graph cut segmentation,” in *ICIP*, 2010.
- [10] G.G. Slabaugh and G. Unal, “Graph cuts segmentation using an elliptical shape prior,” in *ICIP*, 2005.
- [11] P. Das, O. Veksler, Vyacheslav Z., and Y. Boykov, “Semiautomatic segmentation with compact shape prior,” *Image Vision Comput.*, vol. 27, pp. 206–219, January 2009.
- [12] J. Malcolm, Y. Rathi, and A. Tannenbaum, “Graph cut segmentation with nonlinear shape priors,” in *ICIP*, 2007.
- [13] S. Philipp-Foliguet and L. Guigues, “Evaluation de la segmentation d’images : Etat de l’art, nouveaux indices et comparaison,” *Traitement du signal*, vol. 2, pp. 109–124, 2006.
- [14] C. Migniot, P. Bertolino, and J.-M. Chassery, “Contour segment analysis for human silhouette pre-segmentation,” in *VISAPP*, May 2010.