

Segmentation Géodésique Interactive dans le Formalisme du Signal Monogène Couleur

Guillaume DEMARCQ, Hoel LE CAPITAINÉ

Laboratoire Mathématiques, Image et Applications
Université de La Rochelle, Av. M. Crépeau 17042 La Rochelle Cedex, France
guillaume.demarcq01@univ-lr.fr, hoel.le.capitaine@univ-lr.fr

Résumé – Dans cet article, nous présentons un algorithme interactif destiné à la segmentation d’images couleur. L’approche se base sur l’utilisation de points définis par l’utilisateur, qui indiquent les différentes zones à distinguer lors de la segmentation. La segmentation est fondée sur le calcul de distances géodésiques dans le cadre du signal monogène couleur (SMC). Une différence importante par rapport aux précédentes approches, qui utilisent de nombreuses interactions utilisateurs, est que la méthode proposée ne nécessite qu’un nombre très réduit d’interventions. Cela conduit à une utilisation bien plus aisée pour l’utilisateur. Des résultats expérimentaux ainsi que des comparaisons avec des méthodes récentes viennent conforter les bénéfices de notre approche.

Abstract – In this paper, we present an interactive algorithm for segmentation of color images. The user first draws some scribbles into regions that must be discriminated, and the segmentation is then automatically obtained. The segmentation is based on the computation of geodesic distances within color monogenic signal (CMS) fields. An important difference with state-of-the-art methods is that scribbles, which are often segments, are sample pixels picked up by the user. It results in a much more user-friendly segmentation process. Experimental results and comparisons with recent methods show the effectiveness of the approach.

1 Introduction

La segmentation interactive d’images est devenue très populaire ces dernières années, et de nombreuses approches ont été proposées pour aborder ce problème [2, 7, 8]. L’intérêt croissant pour la segmentation interactive s’explique principalement par la grande subjectivité d’une segmentation. Selon l’utilisateur et selon l’image, de nombreuses vérités-terrains peuvent être produites, certaines étant radicalement différentes. La segmentation interactive offre la possibilité à l’utilisateur d’ajouter sa propre information concernant ses désirs sur les objets à distinguer. Une application particulière de la segmentation interactive, sujet de ce travail, est la séparation entre fond et forme dans une image. Dans cet article, l’approche proposée se fonde sur l’utilisation de points définis par l’utilisateur, qui indiquent les différentes zones à distinguer lors de la segmentation. La plupart des approches de ce type nécessitent des segments de grande taille, s’étalant sur l’ensemble de la forme. Ce besoin de décrire de manière complète la forme désirée ne facilite pas la tâche de l’utilisateur. La méthode proposée n’a pour sa part besoin que d’un ensemble réduit de points de référence obtenus par simple clic sur l’image, comme illustré en Figure 1.

L’article est organisé de la façon suivante. En Section 2, après une description des distances géodésiques, de bref rappels sur le signal monogène couleur et ses propriétés sont donnés. Enfin, la méthode de segmentation interactive utilisant le signal monogène couleur est décrite. Des résultats expérimentaux et des comparaisons sont donnés en Section 3, et quelques



FIGURE 1 – Exemple d’interaction pour distinguer forme/fond, et résultats obtenus avec la méthode proposée (gauche : points blanc/rouge) et une méthode récente [2] (droite : traits bleu/vert).

perspectives de travail sont dressées en Section 4.

2 Segmentation Interactive Géodésique

2.1 Distance Géodésique

Soient x et y deux pixels d’une image I à valeurs dans Ω , et $\gamma_{x,y}$ un chemin paramétré par $[0, 1]$ entre ces deux pixels. La

distance géodésique entre \mathbf{x} et \mathbf{y} est définie par

$$d(\mathbf{x}, \mathbf{y}) = \inf_{\gamma_{\mathbf{x}, \mathbf{y}}} \int_0^1 W(\gamma_{\mathbf{x}, \mathbf{y}}(p)) \|\gamma'_{\mathbf{x}, \mathbf{y}}(p)\| dp \quad (1)$$

où $\gamma'_{\mathbf{x}, \mathbf{y}}(p) \in \mathbb{R}^2$ est la dérivée de $\gamma_{\mathbf{x}, \mathbf{y}}(p)$, et W correspond au poids associé à chaque $\gamma_{\mathbf{x}, \mathbf{y}}(p)$. Ainsi, l'utilisation de distance géodésique est essentiellement définie par les poids W utilisés. En pratique, un autre point est d'une importance cruciale, l'efficacité en termes de temps de calcul. Ces distances géodésiques peuvent être évaluées par un algorithme de Fast Marching [10]. Dans cet article, l'implémentation de complexité linéaire proposée dans [11] est utilisée. Le processus de segmentation se fondant sur les distances géodésiques utilise (1) afin de prédire à quelle classe appartient chacun des pixels en sélectionnant la distance minimale avec le fond ou avec la forme. Soit S_l l'ensemble des pixels de référence de forme (F) ou de fond (B). La distance à la référence la plus proche est donnée par

$$D_l(\mathbf{x}) = \min_{s \in S_l} d(s, \mathbf{x}) \quad (2)$$

Enfin, chaque pixel \mathbf{x} est classé comme forme si $D_F < D_B$, en fond sinon.

2.2 Utilisation du Signal Monogène Couleur

Le signal monogène couleur [4] est une extension du concept de signal analytique pour les signaux 1D et du signal monogène pour les signaux 2D tels que des images en niveau de gris. C'est une fonction vectorielle qui peut être décomposée de la manière suivante : soit I une image couleur à valeurs dans le système colorimétrique CIE La*b* (noté Ω) telle que $I = (L, a, b) \in \Omega$, alors le signal monogène couleur¹ qui lui est associé est $f_c = (A_1, A_2, A_3, A_4, A_5) \in \mathbb{R}^2 \times \Omega$, avec

$$\begin{aligned} A_1 &= h_p * h_{R_x} * L + h_p * h_{R_x} * a + h_p * h_{R_x} * b; \\ A_2 &= h_p * h_{R_y} * L + h_p * h_{R_y} * a + h_p * h_{R_y} * b; \\ A_3 &= h_p * L; \quad A_4 = h_p * a; \quad A_5 = h_p * b; \end{aligned}$$

Le terme

$$h_p = \frac{s}{2\pi(x^2 + y^2 + s^2)^{3/2}}$$

est le noyau de Poisson 2D qui donne au signal monogène couleur son caractère multi-échelle et les termes

$$h_{R_x} = \frac{x}{2\pi(x^2 + y^2 + s^2)^{3/2}}$$

et

$$h_{R_y} = \frac{y}{2\pi(x^2 + y^2 + s^2)^{3/2}}$$

correspondent au noyau de Riesz dans les deux directions spatiales x et y respectivement. Le signal monogène couleur incorpore à la fois une information de couleur mais aussi des informations de structures via le noyau de Riesz, et ceci de manière multi-échelle. Par la suite, une représentation polaire

1. De manière plus juste, le signal monogène couleur appartient à l'algèbre de Clifford $\mathbb{R}_{5,0}$ et s'écrit $f_c = \sum_{i=1}^5 A_i e_i$.

$\chi : \mathbb{R}^5 \rightarrow \mathbb{R}_{5,0}$ d'une image couleur par rapport à un vecteur de référence $V = (0, 0, u, v, w)$ est introduite via le produit géométrique² $\chi(V) = f_c V$. De cette représentation, une information d'amplitude \mathcal{A}_V et une information d'angle φ_V peuvent être extraites de la manière suivante :

$$\begin{aligned} \mathcal{A}_V &= |\chi(V)| = |f_c V| = \sqrt{\sum A_i^2 \sqrt{u^2 + v^2 + w^2}}; \\ \varphi_V &= \arctan \frac{|\langle f_c V \rangle_2|}{\langle f_c V \rangle_0} \end{aligned}$$

où $\langle f_c V \rangle_0$ et $\langle f_c V \rangle_2$ sont respectivement les parties scalaires et bivectorielles de $\chi(V)$. Cette nouvelle représentation pour les images couleur apporte des propriétés intéressantes qui vont être exploitées dans une application de segmentation interactive se fondant sur des distances géodésiques.

2.3 Calcul des poids W

Dans ce cadre, il faut que la surface sur laquelle est calculée cette distance reflète un maximum la différence entre les zones de l'image à dissocier. C'est pourquoi les poids W du calcul géodésique ont une importance particulière. Afin d'obtenir une carte des poids la plus pertinente possible, le signal monogène couleur est utilisée au lieu du classique mélange de gaussiennes. Tout d'abord, l'utilisateur choisit manuellement n_F couleurs $(c_F^i)_{i=1, n_F}$ correspondant à la forme (S_F) et n_B couleurs $(c_B^j)_{j=1, n_B}$ correspondant au fond (S_B). Pour chaque couleur de la forme (resp. du fond), le barycentre O_F (resp. O_B) de cette couleur avec les couleurs du fond (resp. de la forme) est calculé. Afin d'améliorer le potentiel discriminant de ces centres, une méthode itérative de recherche d'un centre optimal au sens d'un angle maximum d'une couleur s de forme (resp. de fond) par rapport à l'ensemble des couleurs de fond (resp. de forme) est mise en place. En particulier, on recherche le maximum de la combinaison conjonctive des angles

$$\max_{N(O_i)} \left(\prod_{s_j \in S_{k \neq i}} (\overrightarrow{O_i s}, \overrightarrow{O_i s_j}) \right), \quad (3)$$

où $N(O_i)$ est le voisinage de O_i . L'origine se déplace ainsi de proche en proche jusqu'à convergence. Dans la suite de cet article, le voisinage considéré est composé de l'ensemble des éléments distants d'une unité en norme L_1 . Par changement de repère, *i.e.* en se centrant en O_F (resp. O_B), l'amplitude \mathcal{A}_F^i (resp. \mathcal{A}_B^j) et l'angle φ_F^i (resp. φ_B^j) ($i = 1 \dots n_F$ et $j = 1 \dots n_B$) du signal monogène couleur avec pour vecteur de référence $V = c_F^i$ (resp. $V = c_B^j$) sont obtenus. Ceci permet donc de discriminer les couleurs à la fois en angle et en norme dans l'espace colorimétrique Ω . Ces deux informations sont ensuite fusionnées de manière à ce qu'un angle et une norme faibles correspondent à un faible poids sur la carte géodésique. La fusion sur l'ensemble des n_F (resp. n_B) couleurs de référence s'obtient en conservant les informations les plus importantes

2. Le produit géométrique de deux vecteurs dans les algèbres de Clifford est défini dans [4].

(maximum) discriminantes (minimum) sur les n_l couleurs. En pratique, une composante M_l est obtenue pour chaque l , et la carte géodésique est définie de la manière suivante

$$W_l = \frac{m_l}{m_F + m_B}, \quad (4)$$

où $m_l = \min_{l=c_l}(M_l)$, $l = F, B$. Une fois cette carte obtenue, les distances géodésiques sont obtenues en utilisant (1). A la suite de la première segmentation, un raffinement par apport d’information de l’utilisateur est possible. Ceci passe par la définition d’une nouvelle carte W . A partir d’un nouveau point s_l , un nouveau signal, et une nouvelle origine sont calculés, donnant lieu à la composante M_{s_l} . La carte W est ensuite mise à jour de la façon suivante

$$W_l = \frac{\tilde{m}_l}{\tilde{m}_F + \tilde{m}_B}$$

où

$$\begin{cases} \tilde{m}_l &= (1 - \alpha)M_{s_l} + \alpha m_l \\ \tilde{m}_{k \neq l} &= m_{k \neq l} \end{cases}$$

Le terme de pondération α est choisi arbitrairement, et est fixé dans cet article à $\alpha = \frac{1}{N + 0.2}$, où N est le nombre d’interactions de l’utilisateur.

3 Résultats expérimentaux

3.1 Protocole

Pour évaluer notre approche, nous avons utilisé les 151 images de [7], *i.e.* 49 images de la base de donnée GrabCut [1], 99 images du challenge PASCAL VOC’09 [5], et 3 images provenant de [9]. Comparer et évaluer des méthodes de segmentation interactive n’est pas une tâche évidente car cela peut être fait de manière qualitative ou basé sur un nombre fixe de labels [1]. Cependant cette dernière suppose que l’utilisateur ait déjà décrit la forme de l’objet avec une aire autour des contours de l’objet. Par conséquent, cela n’évalue pas le but principal d’une segmentation interactive basée sur les choix arbitraires de l’utilisateur. Dans [7], les auteurs proposent une évaluation plus adéquate en prenant en compte l’effort (*i.e.* le nombre d’interactions de l’utilisateur) requis pour segmenter une image. Cependant, les différentes méthodes sont évaluées dans un intervalle $A = [85, 98]$, où le score est défini par $O = \frac{c \cap c_{gt}}{c \cup c_{gt}}$, c étant le résultat de la segmentation et c_{gt} la vérité-terrain. Les mauvaises performances avec un nombre restreint d’interactions ne sont donc pas pénalisées par cette mesure d’évaluation. Dans ce papier, nous adoptons une approche un peu différente en considérant la performance moyenne obtenue à partir de la première interaction. Ce choix est fait car un utilisateur peut préférer obtenir des résultats (même avec un score en dessous de 85) avec peu de cliques et ainsi éviter de rajouter beaucoup d’autres points qui n’amélioreraient que très légèrement le résultat. Afin de prendre en compte toutes les interactions

Méthode	AUC normalisée	
	$N = 10$	$N = 20$
GSCseq [7]	83.75%	89.89%
Segmentation géodésique [2]	79.70%	85.65%
BJ [3]	70.94%	78.45%
SMC Segmentation géodésique	84.31%	89.61%

TABLE 1 – Aires sous les courbes (AUC) pour les dix et vingt premières interactions.

de l’utilisateur, nous calculons l’aire normalisée sous la courbe (AUC)

$$AUC = \frac{1}{N} \int_0^N O(t) dt$$

où N est le nombre de label. En raison de la normalisation, le meilleur score en terme d’AUC est celui qui s’approche le plus de 1 (par soucis de clarté, les scores sont donnés sous forme de pourcentages dans le tableau 1).

3.2 Résultats et comparaisons

Nous comparons ensuite nos résultats avec deux méthodes récentes. La première méthode (GSCseq, [7]) est fondée sur la combinaison de contraintes convexes et la formulation de minimisation d’énergie de Boykov-Jolly [3]. Les résultats obtenus par les auteurs sont meilleurs que ceux obtenus par une méthode classique de marche aléatoire [6], si bien que l’on ne reporte que ceux-ci. La seconde approche [2] est purement géodésique, et donc plus proche de nos travaux dans la mesure où aucune contrainte de forme n’est imposée. A ce titre, la comparaison entre les bénéfices de chaque approche sera importante. Nous utilisons le script fourni dans [7] afin d’évaluer l’effort requis par un utilisateur pour obtenir une segmentation correcte. L’évaluation débute avec 3 segments de fond et 1 segment de forme. Puisque notre approche se fonde sur des points, cette évaluation n’est pas adaptée, et nous utilisons la méthode des k -means pour définir les points initiaux. Le nombre d’interventions est fixé à un nombre raisonnable ($N = 10$), et un nombre élevé ($N = 20$). Les aires sous les courbes des 151 images pour GSCseq, GS, BJ et la méthode proposée sont données en Tableau 1. Comme on peut le remarquer, la méthode proposée est plus performante lorsqu’un nombre réduit d’interventions est considéré, tandis que GSCseq est meilleure lorsque N grandit, bien que la méthode SMC reste compétitive. Comme attendu, les résultats sont bons pour un faible nombre d’interactions, et ce sans contraintes de formes. En considérant l’approche purement géodésique, étroitement liée à notre approche, on peut voir que la méthode proposée est clairement meilleure quel que soit le nombre d’interventions. Ceci est principalement lié au cadre de représentation de l’image et du calcul des géodésiques. Des résultats détaillés des courbes de chevauchement sont donnés en Figure 3.

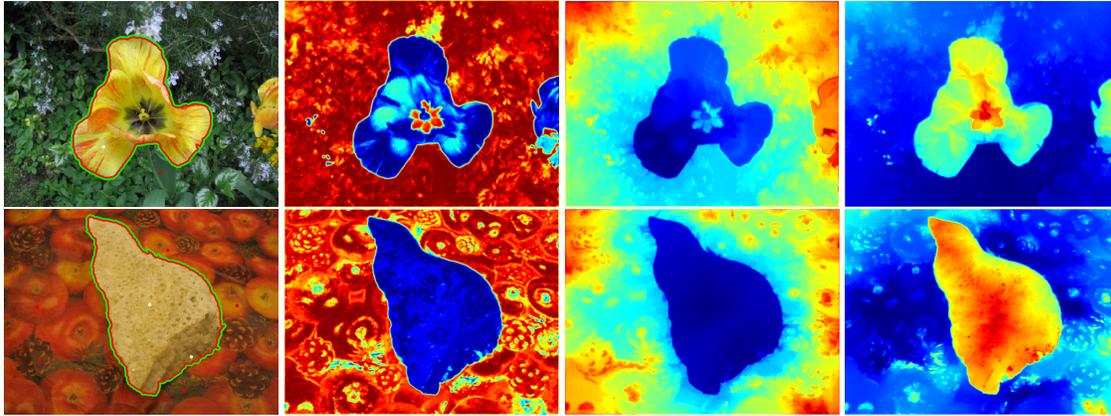


FIGURE 2 – De gauche à droite : images segmentées avec les points, carte des poids W_F , distances géodésiques D_F et D_B .

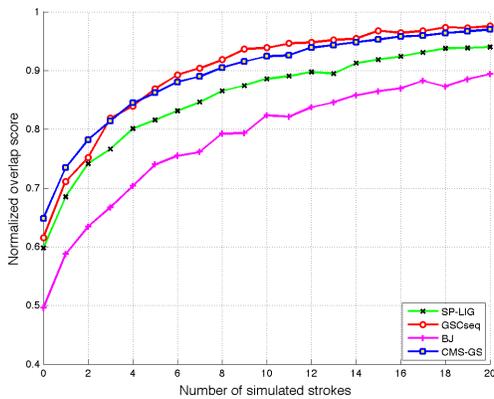


FIGURE 3 – Mesures de chevauchement normalisées en fonction du nombre d'interactions.

4 Conclusion

Dans cet article, une méthode originale de segmentation interactive est proposée. Celle-ci se fonde sur l'utilisation du signal monogène couleur d'une image, et de distances géodésiques. Un apport important se situe au niveau de l'intervention de l'utilisateur ; celui-ci doit seulement cliquer dans les zones désirées, sans tracer de grands traits comme c'est le cas pour les méthodes classiques. Les résultats obtenus, tant au niveau qualitatif qu'au temps de calcul, viennent conforter la démarche proposée. De plus, l'utilisateur peut ajouter des points supplémentaires à l'issue de l'exécution de l'algorithme afin de corriger certains détails, comme c'est le cas pour [2] par exemple. Différentes perspectives peuvent être évoquées à la suite de ce travail : segmentation interactive de vidéos, incrustation géodésique de forme en fonction de l'arrière-plan.

Remerciements

Nous tenons à remercier la région Poitou-Charentes et l'ONR via le NOR Grant N00014-09-1-0493 pour leur soutien financier à ce travail.

Références

- [1] Grabcut image dataset. <http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm>.
- [2] X. Bai and G. Sapiro. Geodesic matting : A framework for fast interactive image and video segmentation and matting. *International Journal of Computer Vision*, 82(2) :113–132, 2009.
- [3] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *Proc. ICCV*, pages 105–112, 2001.
- [4] G. Demarcq, L. Mascarilla, M. Berthier, and P. Courtellemont. The color monogenic signal : Application to color edge detection and color optical flow. *Journal of Mathematical Imaging and Vision*, 40(3) :269–284, 2011.
- [5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge 2009 (voc2009) results.
- [6] L. Grady. Random walks for image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(11) :1768–1783, 2006.
- [7] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *Proc. CVPR*, pages 3129–3136, 2010.
- [8] A. Levin, Rav-Acha, and D. Lischinski. Spectral matting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(10) :1699–1712, 2008.
- [9] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott. A perceptually motivated online benchmark for image matting. In *Proc. CVPR*, pages 1826–1833, 2009.
- [10] J. A. Sethian. A marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.*, 93(4) :1591–1595, 1996.
- [11] L. Yatziv, A. Bartesaghi, and G. Sapiro. A fast $o(n)$ implementation of the fast marching algorithm. *Journal of Computational Physics*, 212(2) :393–399, 2006.