Geodesic interactive segmentation in the color monogenic signal framework

Guillaume Demarcq, Hoel Le Capitaine, Michel Berthier

GEODESIC INTERACTIVE SEGMENTATION IN THE COLOR MONOGENIC SIGNAL FRAMEWORK

G. Demarcq

Explora Nova,
Bordeaux,
France

H. Le Capitaine

LINA UMR CNRS 6241,
École Polytechnique de Nantes,
F-44306 Nantes, France

M. Berthier

Univ La Rochelle,
MIA EA3165,
F-17000 La Rochelle, France

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.

Index Terms— Interactive Segmentation, Clifford Algebra, Color Monogenic Signal, Geodesic Distances

1. INTRODUCTION

Interactive segmentation of images have become very popular in recent years, and many different approaches have been proposed. The growing interest in interactive segmentation is mainly due to the high subjectivity of the process of segmentation. Depending on the user, on the image, the desired result may drastically differ. Interactive segmentation offers the appealing property of allowing a user to give additional information regarding the object of interest. One of the potential applications is to extract an object (foreground) from the background, although the number of classes is not limited to two.

In this paper, the approach is based on the use of points drawn on the image by the user. However, contrary to the vast majority of state-of-the-art approaches [1, 2, 3], the method does not need segments of pixels. Indeed, selecting sample points is generally enough to obtain a convenient segmentation, as illustrated in Figure 1, where two points are sufficient to obtain the segmentation. Our approach is based on a new representation of color images through the color monogenic signal (CMS), which carry both color and structure information. With the help of angular and norm values of this signal, a weight for each pixel is built, and is used to compute geodesic distances. The comparison of distances allows to segment the image into two regions, the foreground and the background.

Fig. 1. User scribbles for background/foreground discrimination with the proposed method (left: white/red points) and a recent method [1] (right: blue/green segments).

2. INTERACTIVE GEODESIC SEGMENTATION AND COLOR MONOGENIC SIGNAL

2.1. Geodesic distance

Let $\mathbf{x}$ and $\mathbf{y}$ be two pixels of an image I lying in Ω, and $\gamma(\mathbf{x}, \mathbf{y})$ a parameterized path between the two pixels. Then, the geodesic distance between $\mathbf{x}$ and $\mathbf{y}$ is given by

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

where $\gamma'_{\mathbf{x},\mathbf{y}}(p) \in \mathbb{R}^2$ is the derivative of $\gamma_{\mathbf{x},\mathbf{y}}(p)$, and W defines the weight for each $\gamma_{\mathbf{x},\mathbf{y}}(p)$. Consequently, when one uses geodesic distance, the most important step is to efficiently set the weights. However, in practice, another issue has to be considered: the computational efficiency. In this paper, we follow the approach of [1], that uses the approach proposed in [4] to compute (1) in linear time. It is based on the resolution of the non-linear Eikonal equation, which is a particular case of the Hamilton-Jacobi equations where H

is reduced to $\|\nabla u(\mathbf{x})\| - F(\mathbf{x})$, $\mathbf{x} \in \Omega$, where $u(\mathbf{x})$ is the shortest time needed to travel from the boundary of Ω to $\mathbf{x}$ with time cost $F(\mathbf{x})$. A first numerical solution was provided in [5], called here after under the generic term *fast marching methods*. The complexity of this approach, as well as some other methods subsequently proposed is $O(N\log(N))$, N being the number of grid points. These techniques are based on upwind numerical schemes. In this paper, we use a novel implementation of the fast marching algorithm proposed in [4], based on untidy priority queues, which provides a solution in linear time ($O(N)$ complexity). Therefore, it is useable in an interactive application such as semi-automatic segmentation. Geodesic segmentation uses (1) to label pixels by selecting the minimum distance with the foreground or the background. Let S_l the set of seeds with label foreground (F) or background (B). The distance to the closest seed of each label is computed by

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

Finally, the pixel $\mathbf{x}$ is labelled as foreground if $D_F < D_B$, and background otherwise.

2.2. Color Monogenic Signal (CMS)

In [6], Demarcq et. al introduce a novel representation for color images called the color monogenic signal. Based on the construction of the analytic and monogenic signal in Clifford algebras, the authors use the Dirac equation in the Clifford algebra $\mathbb{R}_{5,0}$ as a generalization of the Cauchy-Riemann equations in order to construct the CMS. A brief description of basic concepts is given in the following section. Given an image I in a color space Ω, say *e.g.* the CIE Lab color space, such that $I(x,y) = (L(x,y), a(x,y), b(x,y))$, the color monogenic signal f is a vector of $\mathbb{R}_{5,0}$ and reads as following

$$f(x,y,\tau) = A_1e_1 + A_2e_2 + A_3e_3 + A_4e_4 + A_5e_5$$

$$\text{with} \begin{cases} A_1 = h_P * h_R^x * a(x,y) + h_P * h_R^x * b(x,y) \\ \qquad + h_P * h_R^x * L(x,y) \\ A_2 = h_P * h_R^y * a(x,y) + h_P * h_R^y * b(x,y)) \\ \qquad + h_P * h_R^y * L(x,y) \\ A_3 = h_P * a(x,y) \\ A_4 = h_P * b(x,y) \\ A_5 = h_P * L(x,y) \end{cases}$$

$$\text{and} \begin{cases} h_P(x,y,\tau) = \dfrac{\tau}{2\pi(x^2+y^2+\tau^2)^{3/2}} \\ h_R^x(x,y) = \dfrac{x}{2\pi(x^2+y^2)^{3/2}} \\ h_R^y(x,y) = \dfrac{y}{2\pi(x^2+y^2)^{3/2}} \end{cases}$$

The term h_P is the 2D Poisson kernel and acts as low-pass filter on the image. The term h_R is the Riesz kernel and gives an information about the structure in the image. As the reader can see, the color monogenic signal carries both color and structure information in a multiscale way, but it holds other properties as well. The color monogenic signal can be written in a polar form with respect to a unitary chosen vector V and is given by the so-called geometric product

$$\chi(V) = fV$$

The norm of $\chi(V)$ gives the amplitude $\mathcal{A}_V$ of the CMS

$$\mathcal{A}_{_V} = |\chi(V)| = |fV| = |f||V| = |f|$$

and the angle between f and V is given by $\varphi_{_V} = \text{atan}\dfrac{|\langle fV\rangle_2|}{\langle fV\rangle_0}$ where $\langle fV\rangle_0$ and $\langle fV\rangle_2$ denote the scalar and bivector part respectively (see [6] for more details). Finally, the polar representation of the CMS is obtained by $\chi(V) = \mathcal{A}_{_V} e^{B\varphi_{_V}}$, where $B = \langle fV\rangle_2 / |\langle fV\rangle_2|$ is a unitary bivector which squares to minus one and thus can be assimilated to the complex number i.

2.3. Setting the geodesic map W

Setting properly the basis of the computed color monogenic signal is a very important step. Depending on the origin of the colorimetric space (CIE Lab in our method), two similar colors may not be distinguished although the user specifies that one belongs to the foreground and the other to the background. Consequently, we propose to translate the origin by a local optimization procedure. More precisely, for each seed point $s \in S_l$ we compute an origin as follows. The initial position of the origin O_l is given by the weighted barycenter of the seed points belonging to S_l and one of the seed points s_j that does not belong to S_l. Then, the angles between the seed s_j and all the seeds of S_l are computed and combined through a conjunctive operation (here the product):

$$\max_{N(O_l)} \left(\prod_{s \in S_l} (\overrightarrow{O_l s}, \overrightarrow{O_l s_j}) \right), \tag{3}$$

where $N(O_l)$ is the neighborhood of the origin O_l. The origin moves into the direction of maximum angle. This step is repeated until convergence, *i.e.* there are no greater values in the neighborhood of the origin. Here, the neighborhood is 3-dimensional, and is composed of all points at unit L_1 distance in $\mathbb{R}^3$. Once the different origins are obtained, one can compute their representation through the CMS using $\chi(V)$. Therefore, for each seed, there is a corresponding CMS representation characterized by its angles and its norms.

More formally, given n_F seeds f_i for the foreground Ω_F and n_B seeds b_j for the background Ω_B, a new origin O_i between each f_i (resp. b_j) and $(b_1, ..., b_{n_B})$ (resp. $(f_1, ..., f_{n_F})$) is obtained. In each new coordinate system with origin $O_i = (c_1, c_2, c_3)$, a color $c = (a, b, L)$ becomes $\tilde{c} = (a - c_1, b - c_3, L - c_3)$ and the color monogenic signal is computed in this setting. The next step is devoted to use the properties of

the polar representation $\chi(V)$ with $V = f_i$, *i.e.* the amplitude $\mathcal{A}_V$ and the angle φ_V are used. In order to exploit these information for finding pixels with same color and structure, several functions are introduced. For each pixel in the image, $D_V = |\mathcal{A}_V - |V||$ measures the length difference between the two norms $\mathcal{A}_V$ and $|V|$. A normalized field for angle representation is also obtained with $\Phi_V = 1 - \cos\varphi_V$, which is the angular distance to the reference vector V. Afterwards, the two measures are combined in a disjunctive way in order to obtain a small response only when both angular distances and length differences are small. In this paper, we use the probabilistic sum $g(x, y) = x + y - xy$. The combined information is carried in $M_V = g(D_V, \Phi_V)$. This scheme is repeated for each f_i and finally, the minimum over all f_i is taken

$$m_F = \min_{V=f_i}(M_V)$$

Then, the same process is applied with background seeds, and we obtain

$$m_B = \min_{V=b_j}(M_V)$$

These two functions are used to define the foreground weighted map in the geodesic method as $W_F = \frac{m_F}{m_F + m_B}$. Once the minimum paths for each pixel and each seed are obtained, the classification is made by using (2) for the foreground and the background. Following the first segmentation, a refinement is possible by adding information coming from the user. To do so, the map W must be modified. Starting from a new point s_l, a new signal, and a new origin are computed, giving a new component M_{s_l}. The map W is then updated as follows

$$W_l = \frac{\tilde{m}_l}{\tilde{m}_F + \tilde{m}_B}, \quad \text{where} \begin{cases} \tilde{m}_l & = (1-\alpha)M_{s_l} + \alpha m_l \\ \tilde{m}_{k\neq l} & = m_{k\neq l} \end{cases}$$

The weighting term α is arbitrary chosen, and is set in this paper to $\alpha = 1/(N + 0.2)$, where N is the number of interactions of the user.

3. EXPERIMENTAL RESULTS

3.1. Protocole

In order to evaluate the approach, we use the 151 images as in [2], *i.e.* 49 images from the GrabCut dataset [7], 99 images from the PASCAL VOC'09 segmentation challenge [8], and 3 images coming from the alpha-matting repository [9]. Evaluating interactive segmentation is a difficult task. It may be a qualitative opinion as in [1], or based on a fixed set of seeds [7]. However, the latter supposes that the user has already described the shape of the object with an uncertain area around the edges of the objects. Therefore it does not evaluate the real objective of interactive segmentation based on scribbles. In [2], the authors proposed a more convenient evaluation criteria by evaluating the effort (*i.e.* the number of user interactions) required to segment an image. However, the methods are evaluated in a interval of overlap scores $A = [85, 98]$, where the overlap score is defined by $O = \frac{c \cap c_{gt}}{c \cup c_{gt}}$, c being the segmentation result and c_{gt} the ground truth. Therefore, poor performances with a low number of user interaction is not penalized by their evaluation measure. In this paper, we adopt a slight variation of their quality evaluation by considering the mean performance obtained starting from the first user interaction. This setting is used because a user might prefer to obtain results (even with an overlap score lower than 85) with a few number of points, instead of adding many points for a small improvement. Therefore, this evaluation is not adapted to point-based segmentation methods, but we adopt it for comparison purpose. In order to take into account all user interactions, we compute the normalized area under the curve (AUC)

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

where N is the number of strokes, and t denotes the user interactions. Due to the normalization, the best achievable AUC is 100%, and the larger the better.

3.2. Results and comparisons

First, we give in Figure 2 two original images with their weight maps, their D_F maps. We compare our results with

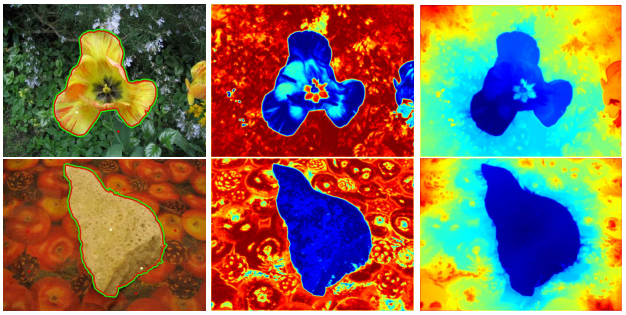

Fig. 2. From left to right: segmented images with the strokes, weight map W and geodesic distances D_F.

two recent methods. The first one (GSCseq, [2]) is based on the combination of star-convexity constraints with the Boykov Jolly (BJ) energy formulation [10]. Results reported by the authors outperform the results obtained by Random Walks [11], so that we focus on the GSC method and its initial approach BJ. The second one (SP-LIG, [1]) is purely based on geodesics without prior on shapes, like the proposed method. Therefore, our work is more related to this method, and a special study of their relative performance is carried on. We use the robot script from [2] to evaluate the effort of the user to obtain a segmentation. Originally, the evaluation starts with 3 background scribbles and 1 foreground scribble that have been manually selected. However, since our approach relies on points, the evaluation is not well adapted. Consequently, samples points are selected through a k-means

Method	Normalized AUC	
	$N = 10$	$N = 20$
GSCseq [2]	83.75%	**89.89%**
Geodesic Segmentation (SP-LIG)[1]	79.70%	85.65%
BJ [10]	70.94%	78.45%
CMS Geodesic Segmentation	**84.31%**	89.61%

Table 1. Area under the curves for the ten and twenty first user interactions.

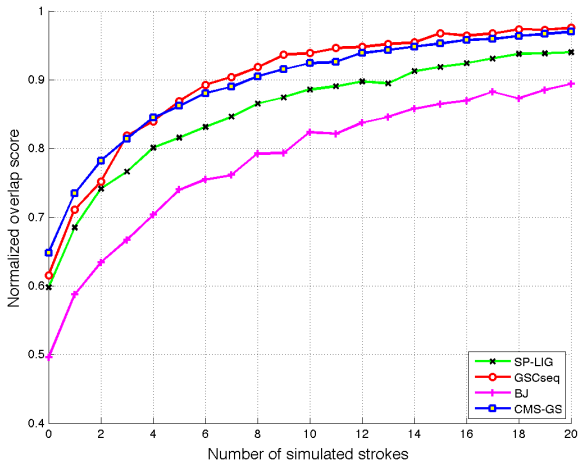

Fig. 3. Normalized overlap scores as a function of the number of strokes.

algorithm in the scribbles to define the seeds. In the experiments, a reasonable number $N = 10$ user interactions and a larger number $N = 20$ user interactions are considered. We give in Table 1 the area under the curves obtained on the 151 images with GSCseq, Geodesic Segmentation (SP-LIG) and BJ methods. As can be seen, the proposed method is better than all other methods when $N = 10$ strokes are considered, while GSCseq performs better when the user has drawn $N = 20$ scribbles. The results confirm that our method allows to obtain satisfying results with a few number of points. Additionally, it does not include any prior on the shape of the objects to be extracted. Although our method provides a lower score in the case of $N = 20$, it remains competitive with GSCseq. Consequently, adding shape priors such as star-convexity in GSCseq would increase the performance on the whole dataset. Considering the purely geodesic approach [1], closely related to our approach, one can see that the CMS geodesic segmentation allows to obtain better results. This is mainly due to the new representation of color images given by the CMS. Detailed overlap curves for all methods are plotted in Figure 3.

4. CONCLUSION

In this paper, a new method of interactive segmentation of color images is presented. This method is based on the dual use of the color monogenic signal representation of color images and geodesic distances. The main advantage of our contribution is the number of strokes needed for segmentation, therefore giving an easy and intuitive tool for interactive segmentation. Although our method does not outperform the GSC method for a large number of strokes, it remains competitive with it, and outperforms older methods. The method has the appealing property of obtaining satisfying results with a few number of points. For the end-user, this property may be crucial when choosing his processing software. Among the perspectives we have in mind, we want to study the benefits of adding star-convex shapes prior [2] into the proposed approach. Additionally, we want to work on a method allowing realistic composition of two images, where luminance information of the background is taken into account for foreground inlaying.

5. REFERENCES

[1] X. Bai and G. Sapiro, "Geodesic matting: A framework for fast interactive image and video segmentation and matting," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 113–132, 2009.

[2] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *Proc. CVPR*, 2010, pp. 3129–3136.

[3] A. Levin, Rav-Acha, and D. Lischinski, "Spectral matting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1699–1712, 2008.

[4] L. Yatziv, A. Bartesaghi, and G. Sapiro, "A fast o(n) implementation of the fast marching algorithm," *Journal of Computational Physics*, vol. 212, no. 2, pp. 393–399, 2006.

[5] J. A. Sethian, "A marching level set method for monotonically advancing fronts," *Proc. Nat. Acad. Sci.*, vol. 93, no. 4, pp. 1591–1595, 1996.

[6] 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*, vol. 40, no. 3, pp. 269–284, 2011.

[7] "Grabcut image dataset," http://research.microsoft.com/en-us/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge 2009 (voc2009) results.," .

[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*, 2009, pp. 1826–1833.

[10] Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *Proc. ICCV*, 2001, pp. 105–112.

[11] L. Grady, "Random walks for image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.