MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Motion Segmentation Using Divisive Graph Cuts

Jong-Sung Kim, Il-Kwon Jeong
Electronics and Telecommunications Research Institute (ETRI), Korea
{js.kim, jik}@etri.re.kr

Abstract

In this paper, we present a graph cut-based motion segmentation method that takes occlusion into account. We formulate the motion segmentation problem in terms of energy minimization with accounting for occlusion and minimize the energy function with the divisive graph cut algorithm where multiway minimum cuts for motion segmentation are efficiently computed through the swap move and split move of binary labels. A graph cut-based motion estimation technique is employed to estimate the motion field and occlusion between consecutive frames of the motion image sequence. Based on the motion estimate, our method segments a current frame into a number of regions of similar motion by assigning a label to each pixel. The label assignment of occluded pixels, of which the motion is not defined, is determined based on a color prior. The effectiveness of our method was verified with experimental results for various real motion image sequences.

1 Introduction

The aim of this work is to investigate an optimal motion segmentation method for motion image sequences with occlusion, of which an example is shown in Fig. 1. Although the motion estimation and segmentation has been extensively studied, the motion estimation as well as the motion segmentation still remains a challenging problem because occlusion occurs commonly in real image sequences. Graph cut methods [1, 2] showed very promising motion estimation performance, with approximately 4 times fewer errors than standard methods such as normalized correlation. Besides, graph cut methods have important advantages which other ones do not have. For example, the graph cut method proposed by Kolmogorov and Zabih [3], referred by the KZ method in this work, can addresses occlusions properly, while preserving other advantages of graph cuts, such as optimality. However, the motion segmentation based on such motion estimate is still error-prone due to the motion estimation error around occluded areas as illustrated in Fig. 1 where we can see that occlusion occurs around motion boundary and the motion of occluded pixels is not defined.

The motion segmentation problem of this work is formulated in terms of energy minimization, which results in a new motion segmentation energy function based on the motion field and occlusion between two consecutive frames of an input motion image sequence. The KZ method is employed to estimate the motion field and occlusion. An additional variable is introduced to account for the occlusion in our formulation. Then, we efficiently minimize the energy function with

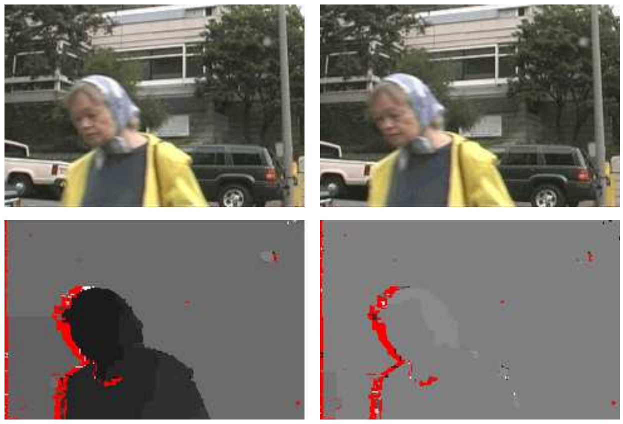

Figure 1: A motion image sequence and its motion field and occlusion estimation result with the KZ method [3]. Red-colored areas denote occlusion.

divisive graph cuts (DGC), which has been introduced in [4] to compute multi-way minimum cuts through the swap move and split move without user-defined seeds contrary to other graph cut methods [5, 6, 7]. Using the DGC algorithm, we optimally assign a label to each pixel according to the motion estimate to segment a current frame into a number of regions of similar motion. A color prior is used to solve the label assignment of occluded pixels by enforcing the label of each pixel likely to be the same with neighboring pixels of similar color.

The main advantage of our method is that it can find an optimal motion segmentation for a motion image sequence from its inaccurate motion estimate with occlusions. In the previous segmentation methods, the segmentation of motion is tackled by applying K-means clustering [8], Maximum-Likelihood framework [9], or Normalized cuts [10]. However, those methods belong to a sub-optimal solution from a computational point of view. Besides, they cannot handle occlusions properly contrary to ours. In addition, our method is different from layer-based motion segmentation methods using graph cuts [11, 12], where a motion image sequence is represented with a number of layers of specific motion, such as affine motion or projective one. On the contrary, our method does not depend on such image motion model. The effectiveness of our method was verified with experimental results for various real motion image sequences.

This paper is organized as follows. Section 2 explains the formulation and algorithm of our method in details. In Section 3, we give experimental results of our method for real image sequences. Our concluding remarks are presented in Section 4.

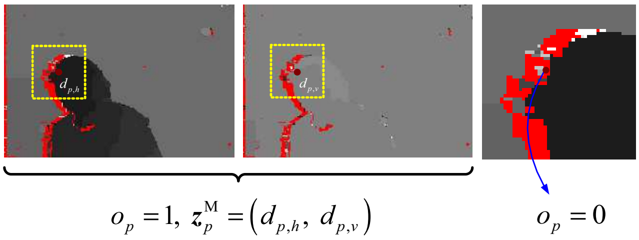

Figure 2: An illustration of the data term with the motion vector, $\boldsymbol{z}_p^{\mathrm{M}}$, and the occlusion variable, o_p, for a pixel, p, on a current frame of *Woman* sequence.

2 Proposed Method

2.1 Motion Segmentation Energy

Let us denote I and J two consecutive frames from a motion image sequence. The motion field between I and J is denoted by $\boldsymbol{z}^{\mathrm{M}} \equiv \{\boldsymbol{z}_p^{\mathrm{M}}\}_{p=1}^{|\mathcal{P}|}$, where $\boldsymbol{z}_p^{\mathrm{M}}$ is the motion vector at some pixel p in I and $\mathcal{P}$ is the set of pixels in I. The motion vector is defined as $\boldsymbol{z}_p^{\mathrm{M}} \equiv (d_{p,h}, d_{p,v})^\top$, where $d_{p,h}$ and $d_{p,v}$ are the horizontal and vertical disparity values. In this work, the motion vector $\boldsymbol{z}^{\mathrm{M}}$ is estimated with the KZ method [3]. Then, an additional variable $\boldsymbol{o} \equiv \{o_p\}_{p=1}^{|\mathcal{P}|}$ is introduced to denote the occlusion or non-occlusion for each pixel p as

$$o_p = \begin{cases} 0 & \text{if } p \in \mathcal{O} \\ 1 & \text{otherwise} \end{cases}, \tag{1}$$

where $\mathcal{O}$ is the set of occluded pixels. Note that the motion vector $\boldsymbol{z}_p^{\mathrm{M}}$ is not defined at occluded pixels as shown in Fig. 2. Therefore, it is a challenge to assign a motion segmentation label to those pixels. To solve this problem, the RGB color of the current frame, denoted by $\boldsymbol{z}^{\mathrm{C}}$, is used in a prior term of the energy.

Let A be a configuration (or labeling) corresponding to the motion segmentation. In our method, the energy function for motion segmentation, E, has the form

$$E(A, \Theta) = \lambda \cdot D(A, \Theta) + K(A), \tag{2}$$

where the data term D measures how the motion of each pixel confirms to the parametric mixture of motion data and the prior term K imposes a penalty if neighboring pixels of similar color have different segmentation labels and the parameter $\lambda > 0$ controls the balance between the two terms. We assume that the motion data $\boldsymbol{z}^{\mathrm{M}}$ are generated from a finite mixture density model $G(\boldsymbol{z}^{\mathrm{M}}|\Theta) = \sum_{l=1}^{L} g_l(\boldsymbol{z}^{\mathrm{M}}|\theta_l)$ with unknown L components, where $\Theta \equiv \{\theta_l\}_{l=1}^{L}$ is a set of model parameters and L is the unknown number of components of the mixture model. By integrating the occlusion variable $\boldsymbol{o}$, the data term is defined by

$$D(A, \Theta) = \sum_{p \in \mathcal{P}} o_p \cdot D_p(A_p, \Theta), \tag{3}$$

which explicitly reflects that the data term should be zero if $o_p = 0$, i.e., the motion data should not produce an effect on the label assignment of occluded pixels. The value of the individual term D_p is defined by

$$D_p(A_p = l, \Theta) = \begin{cases} \frac{1}{2} r_{p,l}^2 & \text{if} \quad |r_{p,l}| \le \zeta_l \\ \zeta_l \cdot \left(|r_{p,l}| - \frac{1}{2}\zeta_l\right) & \text{otherwise} \end{cases} \tag{4}$$

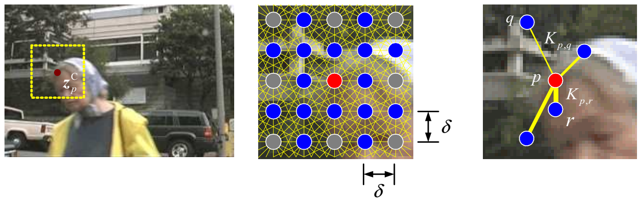

Figure 3: An illustration of prior penalties in a 16-neighborhood system. The strength of penalty is reflected by the thickness of line.

where $r_{p,l}$ is the residual term

$$r_{p,l}^2 = \frac{1}{2}\left(\boldsymbol{z}_p^{\mathrm{M}} - \boldsymbol{m}_l\right)^\top \mathrm{S}_l^{-1}\left(\boldsymbol{z}_p^{\mathrm{M}} - \boldsymbol{m}_l\right). \tag{5}$$

with the mean vector $\boldsymbol{m}_l$ and covariance matrix S_l of the motion vector $\boldsymbol{z}_p^{M}$. The tuning parameter ζ_l is robustly computed by $\zeta_l = \tau \cdot \text{median}\{|r_{p,l}|\}$, with $\tau = 1.4826$ typically, where the value of τ is chosen to make the robust estimate of ζ_l to be consistent with the standard deviation of normal density.

The prior term is defined by

$$K(A) = \sum_{\{p,q\} \in \mathcal{N}} K_{p,q} \cdot \delta(A_p \neq A_q), \tag{6}$$

where

$$\delta(A_p \neq A_q) = \begin{cases} 1 & \text{if} \quad A_p \neq A_q \\ 0 & \text{if} \quad A_p = A_q \end{cases}, \tag{7}$$

represents the Potts interaction model to encourage the labeling A to be piecewise constant and $\mathcal{N}$ denotes a set of all pairs of pixels neighboring in some neighborhood system such as the 16-neighborhood in Fig. 3. For each neighboring pixel pair (p, q) in $\mathcal{N}$, the term $K_{p,q}$ is defined by

$$K_{p,q} = \left(\kappa \cdot \exp\left(-\frac{1}{2} d_{\mathrm{H}}^2\left(\boldsymbol{z}_p^{\mathrm{C}}, \boldsymbol{z}_q^{\mathrm{C}}\right)\right) + 1 - \kappa\right) \cdot \frac{\eta}{d(p,q)}, \tag{8}$$

with the parameter $\kappa \in [0, 1]$ to determine the influence of the exponential function of $d_{\mathrm{H}}\left(\boldsymbol{z}_p^{\mathrm{C}}, \boldsymbol{z}_q^{\mathrm{C}}\right)$, the distance between two color feature vectors $\boldsymbol{z}_p^{\mathrm{C}}$ and $\boldsymbol{z}_q^{\mathrm{C}}$, normalized with the covariance matrix, H, of the quantities $\boldsymbol{z}_p^{\mathrm{C}} - \boldsymbol{z}_q^{\mathrm{C}}$ for all pairs of p and q in $\mathcal{N}$. The second term $\frac{\eta}{d(p,q)}$ is a weighting function approximating the length of curves on a regular grid of image, where the parameter $\eta > 0$ is a constant related to the structure of the given neighborhood system $\mathcal{N}$ on the grid of image and the function $d(p, q)$ is the distance between p and q. In this work, the value of η is set with $\eta = \delta^2 \cdot \pi / \iota$, where δ is the cell size, typically set to one, and ι is the size of neighborhood, e.g., $\iota = 16$ for the 16-neighborhood system in Fig. 3.

Figure 4: An example of the split move (the 2nd image) and swap move (the 3rd image) of DGC in our motion segmentation algorithm.

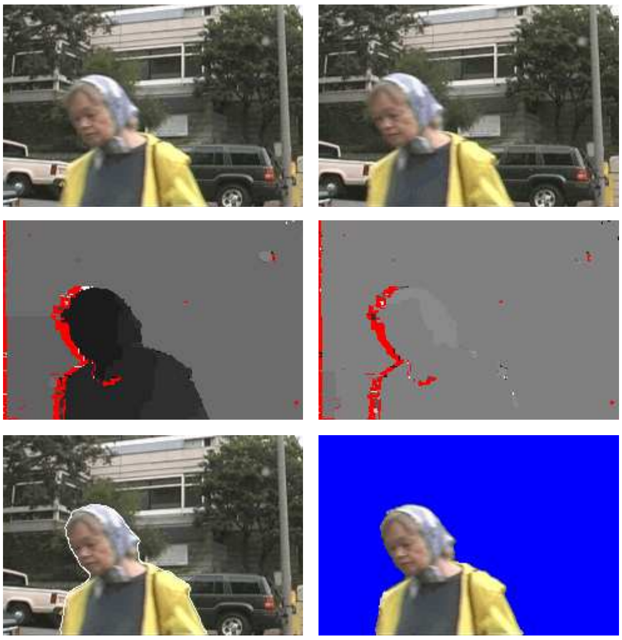

Figure 5: Results of the motion estimation of the KZ method and the motion segmentation of our method for *Woman* sequence.

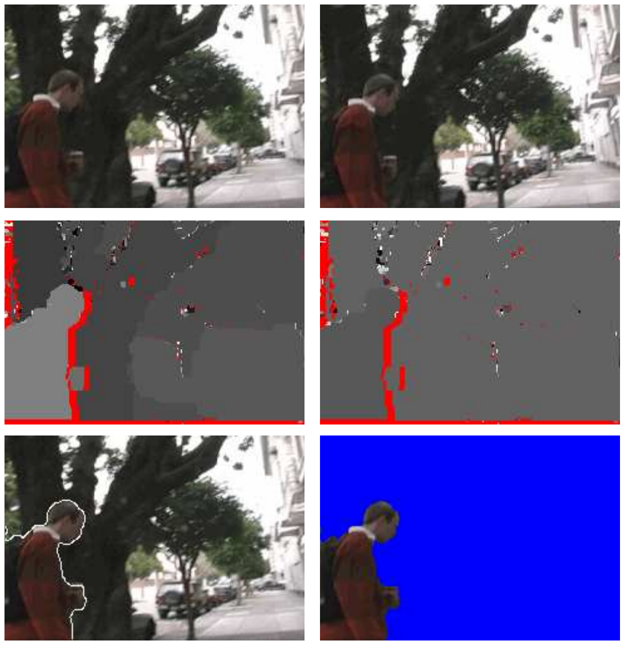

Figure 6: Results of the motion estimation of the KZ method and the motion segmentation of our method for *Man* sequence.

2.2 Motion Segmentation Algorithm

The motion segmentation A can be computed as a global minimum of the motion segmentation energy $E(A, \Theta)$ by solving a multiway cut problem in a weighted graph with multiple terminals. However, solving a multiway cut problem is known to be NP-hard [13]. In addition, the estimation of A and Θ cannot be initiated without some prior knowledge such as seed points or contours [5, 6]. We overcome these difficulties by applying the DGC algorithm [4] which solves the multiway cut problem through the swap move of binary labels in a top-down way and integrates the split move into the swap move to initiate A and Θ without user interaction. The swap move and split move of DGC is recursively applied to a current region to obtain new subregions until the number of motion segments reaches to a pre-defined value or some stopping conditions are met. Fig. 4 illustrates the split move and swap move of DGC in our motion segmentation.

3 Experimental Results

The effectiveness of our graph cut-based motion segmentation method was empirically verified with four kinds of real motion image sequences of 240×160 pixels, *Woman*, *Man*, *Car-People* and *Cars* sequences obtained from the UCLA CIVS Lab (`http://civs.stat.ucla.edu`). The graph cut library [2] was used for algorithm implementation, and a regular grid-fashioned, weighted graph with the 16-neighborhood system was constructed. For explanation, the motion segmentation result of our method for each motion image sequence was presented with two input frames and the motion and occlusion estimation result between the frames. The value of the balancing parameter λ in our method was empirically selected to produce a good segmentation result for each motion image sequence.

Figure 5 shows the motion segmentation result of our method on *Woman* sequence with the motion estimation result of the KZ method. As shown in that figure, the input frame has two dominant motions, i.e., one motion corresponds to the woman and the other corresponds to the background scene. However, the motion estimation result shows that there are many local motions and occluded pixels. Note that the region of woman has at least four kinds of different local motions and many occluded pixels at the left side. However, our method obtained the motion segmentation boundary with a high accuracy. Note that our method could determine the label assignment of occluded pixels without ambiguity. In addition, the number of segments was determined without user interaction. In Fig. 6, we also see that there are also two dominant motions and a variety of local motions, especially at the background region, between the two input frames. Nevertheless, our motion segmentation method found two dominant motions successfully as shown in that figure. In addition, occluded pixels were properly labeled and the segmentation boundaries were accurately localized.

In Figures 7 and 8, our motion segmentation method was applied to *Car-People* and *Cars* sequences, of which each input frame has three dominant motions. As shown in those figures, our motion segmentation method successfully found three dominant motions, while the KZ method gave highly inaccurate motion estimation results for those sequences. As shown in those figures, it was more difficult to obtain accurate segmentation boundaries since the colors of objects in those sequences were very similar to each other. For example, the motion segmentation boundaries in Fig. 7 were not so smooth as true ones, and the segmentation of the larger car in Fig. 8 was leaked to the area of the road due to the ambiguity of color.

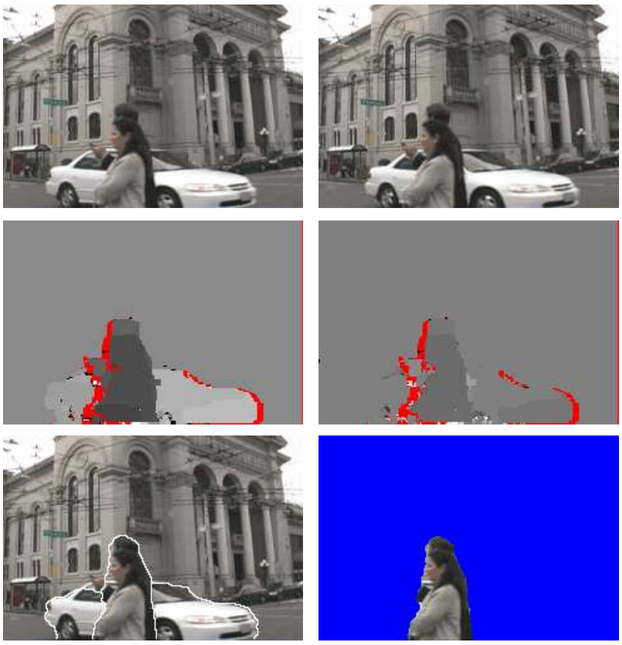

Figure 7: Results of the motion estimation of the KZ method and the motion segmentation of our method for *CarPeople* sequence.

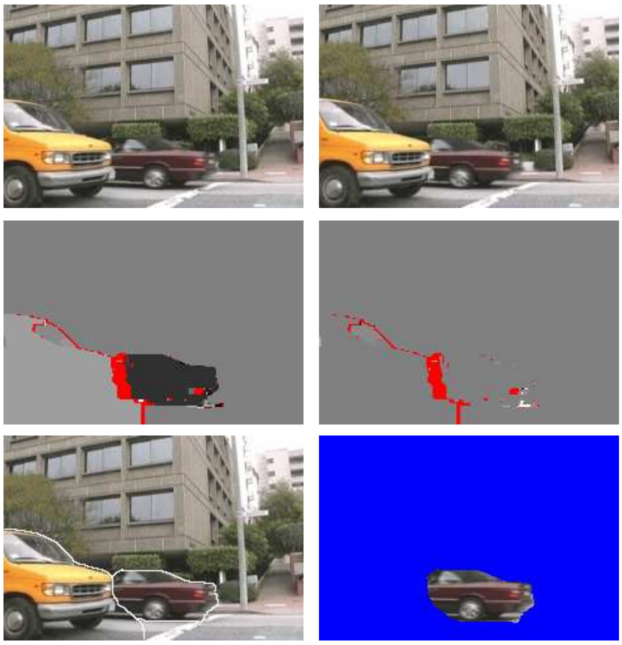

Figure 8: Results of the motion estimation of the KZ method and the motion segmentation of our method for *Cars* sequence.

4 Conclusions

A new divisive graph cut-based motion segmentation method was presented in order to find an optimal motion segmentation of a motion image sequence with occlusion. The motion segmentation problem was formulated with a motion segmentation energy function based on the motion field and occlusion estimate and solved using the divisive graph cut algorithm, which made it possible to compute multiway minimum cuts for optimal motion segmentation without user interaction contrary to other graph cut methods. In this way, our method successfully segmented a current frame into a number of regions of similar motion although the motion field and occlusion estimate was inaccurate. The experimental results of motion segmentation for real motion image sequences showed the effectiveness and performance of our method.

Acknowledgements

This work was supported by the IT R&D joint program of Ministry of Knowledge Economy / Institute for Information Technology Advancement and Ministry of Culture, Sports and Tourism / Korea Culture & Content Agency, [2007-S-051-01, Software development of the Digital Creature].

References

[1] R. Szeliski and R. Zabih: "An Experimental Comparison of Stereo Algorithms," In Proceedings of International Workshop on Vision Algorithms: Theory and Practice, pp. 1-19, 1999.

[2] Y. Boykov, O. Veksler, and R. Zabih: "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11), 1222-1239, 2001.

[3] V. Kolmogorov and R. Zabih: "Computing Visual Correspondence with Occlusions using Graph Cuts," In Proceedings of IEEE International Conference on Computer Vision, pp. 508-515, 2001.

[4] Jong-Sung Kim and Ki-Sang Hong: "A New Graph Cut-based Multiple Active Contour Algorithm without Initial Contours and Seed Points," *Machine Vision and Applications* 19(3), 181-193, 2008.

[5] Y. Boykov and M.-P. Jolly: "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D images," In Proceedings of IEEE International Conference on Computer Vision, pp. 105-112, 2001.

[6] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum: "Lazy Snapping," *ACM Transactions on Graphics* 23(3), 303-308, 2004.

[7] Y. Boykov and G. Funka-Lea: "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision* 70(2), 109-131, 2006.

[8] J. Wang and E. Adelson: "Representing Moving Images with Layers," *IEEE Transactions on Image Processing* 3(5), 625-638, 1994.

[9] S. Khan and M. Shah: "Object based Segmentation of Video using Color, Motion and Spatial Information," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 746-751, 2001.

[10] J. Shi and J. Malik: "Normalized Cuts and Image Processing," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 888-905, 2000.

[11] S. Birchfield and C. Tomasi: "Multiway Cut for Stereo and Motion with Slanted Surfaces," In Proceedings of IEEE International Conference on Computer Vision, pp. 489-495, 1999.

[12] J. Xiao and M. Shah: "Motion Layer Extracting in the Presence of Occlusion using Graph Cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10), 1644-1659, 2005.

[13] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis: "The Complexity of Multiterminal Cuts," *SIAM Journal on Computing* 23(4), 864-894, 1994.