MOTION-DRIVEN OBJECT SEGMENTATION IN SCALE-SPACE

Ebroul Izquierdo M. and Mohammed Ghanbari

Department of Electronic Systems Engineering
University of Essex
Colchester, CO4 3SQ, United Kingdom

ABSTRACT

In this paper we present a method for motion segmentation, in which accurate grouping of pixels undergoing the same motion is targeted. In the presented technique true object edges are first obtained by combining anisotropic diffusion of the original image with edge detection and contour reconstruction in the inherent scale-space. Contours are then matched according to the distance given by a metric defined on their polygonal approximations and the shape of the one-dimensional intensity function along the contour. Masks of objects are obtained by merging image areas inside of edges having the same motion. The performance of the presented technique has been evaluated by computer simulations.

1. INTRODUCTION

The accurate solution to the object segmentation task is crucial in most emerging content-based techniques supporting the next generation of video coding standards (MPEG4 and MPEG7) and multimedia documents. Currently there exists a large collection of application domains for these key technologies including: entertainment, education, medical imaging, augmented reality and immersive telepresence systems. In recent years, great efforts have been made to developed motion-driven methods for object segmentation. Among others, Ibenthal et al. [3] describe a method in which unlike the contour-matching approach described in this paper, a hierarchical segmentation scheme is applied. The motion field is used in order to improve the temporal stability and accuracy of segmentation. Chang et al. [2] introduced a Bayesian framework for simultaneous motion estimation and segmentation based on a representation of the motion field as the sum of a parametric field and a residual field. Borshukov et al. [1] present a multiscale affine motion segmentation based on block affine modeling. Although in all these works the dynamic of the objects present in the scene is used to enhance the segmentation results, the extraction of accurate object masks is not challenged because less attention is paid to the spatial reconstruction of objects contours as basis for object mask determination. In this context, Izquierdo and Kruse [4] describes a method for accurate object segmentation but in contrast to the technique introduced in this paper their approach is tailored for stereoscopic sequences using disparity information and morphological transformations.

In the approach presented in this paper, each frame in the sequence is first processed by applying a non-linear diffusion method in which averaging is inhibited in the direction of relevant edges and smoothing the image in other regions. The goal of this processing step is to enhance edges keeping their correct position, reduce noise and smooth regions whit small intensity variations. This initial processing step is detailed in the next section. Image edges are then extracted at the location where the second derivative of the anisotropic-diffused image crosses zero (zero-crossings of the Laplacian). Small gapes in the resulting contours are then closed by linking them with straight lines. To simplify the contour matching procedure, in a last processing step edges are approximated by polygonal lines. These both aspects (edge extraction and linearization) are described in section 3. The problem of finding the best fit between image edges in two consecutive frames is finally solved by measuring their similarity via a suitable metric defined on the polygonal lines and the shape of the one-dimensional intensity function along the contour. Using the fact that different objects can be completely described by relevant edges and their motion, object masks are extracted by merging image areas undergoing similar motion in connected regions enclosed by a contour. Section 4 deals with this last processing step. Selected results of computer simulations and conclusions are drawn in section 5.

2. EDGE ENHANCEMENT BY ANISOTROPIC DIFFUSION

In the context of non-linear scale space, a set of images $I(x,y,t)$ is generated, with $I(x,y,0)$ as the original image and t as scale parameter, by applying the diffusion equation of porous medium type

$$I_t = \nabla \cdot [c(x,y,t)\nabla I] = div(c(x,y,t)\nabla I)\,, \qquad (1)$$

where the diffusion velocity c depends on the local energy at the image position (x, y). If c is chosen as a function of the image edges, the diffusion process should tend to a piece-wise constant solution representing a simplified image with sharp boundaries. That is, the amount of diffusion in each image point has to be modulated by a function of the image gradient at that point. Furthermore, c has to be a continuously decreasing function of the image gradient, so that image regions of high contrast undergo less diffusion, whereas uniform regions are diffused with the same intensity in all directions. The evolution of the diffusion process described by this nonlinear partial differential equation yields a three dimensional solution space. A cross-section of the surface at a particular time t describes the diffusion result that we are interested in. Using the diffusion equation (1) with $c(x,y,t) = f(\|\nabla I(x,y,t)\|^2)$, a set of

anisotropic-diffused images is generated. Different choices for f are proposed in the literature. In our work we use $f(w) = A/(1+\frac{w}{B})$ as proposed by Perona and Malik [6]. Note that the two parameters A and B, control the amount of diffusion. If $w < B$, then $f(w)$ tends to A, whereas for $w > B$, $f(w)$ tends to 0. Fig. 1 illustrates this diffusion process for the test sequence CAR. More detailed description of these results are given in section 5.

Fig. 1: Diffusion preprocessing for the sequence CAR. Original image (top) and anisotropic-diffused image (bottom).

3. EDGE DETECTION AND LINEARIZATION IN SCALE-SPACE

Main image contours are detected at the location where the second derivative of the anisotropic-diffused image crosses zero (zero-crossings of a cross-section at time t of the solution space of (1)). We have chosen this second order differential operator because it satisfies the above stated properties when applied to noise-free images. Zero-crossings are extracted by regarding a 3X3 neighborhood of each sampling position in the Laplacian-filtered images. Relevant object edges are identified in the image obtained for a large scale-value t. They are then completed using edges extracted from less diffused images. For a given discretization of the scale parameter $t_0,\ldots,t_k,\ldots,t_n$, we start with the diffused image at scale t_n and iteratively passing from one image to the next less diffused image, the whole object edge is reconstructed. As stated by Perona and Malik [6], a generic choice of $c(x,y,t)$ not necessarily guarantees that zero-crossings of the Laplacian satisfy the causality condition. This means, that zero-crossings at $I(x,y,t_n)$ can not exist or appear at different positions in less diffused images Nevertheless, our experiences show that these distortions are sufficiently small to be corrected by applying a simple procedure in which edges detected at large scales are shifted to the position indicated by the same edges at lower scales. Once the edge completion process has been carried out, remaining small gaps in the zero-crossing contours are closed by straight line segments in order to obtain as many close contours as possible. Finally, all the edges whose length do not exceed a given value are removed. Fig. 2 shows the relevant edges extracted from the images presented in Fig. 1.

Before contours in two consecutive frames are matched in order to detect their motion, they are approximated by polygonal lines. A suitable procedure to cope with this task has been described in detail in [5], [7]. The method consists of expanding the considered curve to a narrow δ-band and finding the shortest polygonal path lying in the strip defined by the δ-band. These approach produces a polygonal path for each contour avoiding the worst effects of distortions in the contour. The degree of smoothing is controlled by the bandwidth δ. Fig. 3 shows the polygonal approximation of the curves shown in Fig 2 by applying this technique.

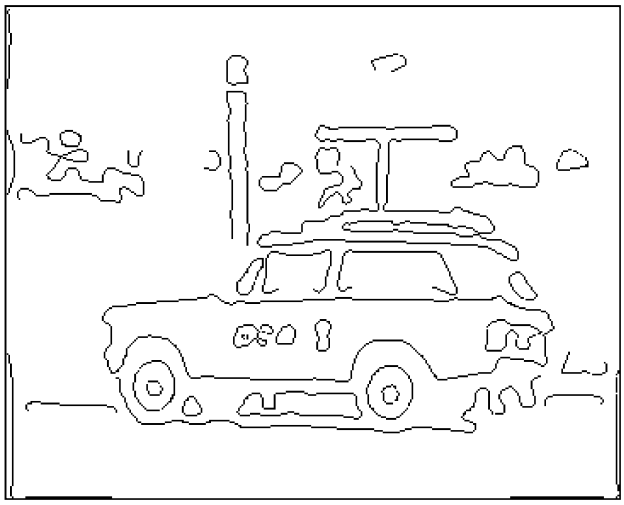

Fig. 2: Relevant edges of the image shown at the top of Fig. 1.

4. CONTOUR MATCHING

To achieve our final segmentation purpose, motion of the contours is estimated by applying a technique based on a metric for contour similarity, which depends directly on the global contour shape and the shape generated by the intensity values along it.

Let $ZC^t = \{C_0^t, C_1^t, \ldots, C_p^t\}$ and $ZC^{t+1} = \{C_0^{t+1}, C_1^{t+1}, \ldots, C_m^{t+1}\}$ be the sets of all detected relevant edges in the image at time *t* and *t+1,* respectively. For a given contour $C_i^t \in ZC^t$ our next task is to find the contour in ZC^{t+1} that best fits with C_i^t. To solve this, we consider the corresponding polygonal lines. The matching is carried out using an appropriate metric for comparing polygonal shapes. The contour corresponding to the polygonal line in ZC^{t+1} with the shortest distance to C_i^t is then chosen as the best match.

A natural way to define a metric for comparing polygonal paths is to use the distance between their turning functions. The turning function $\Theta_C(s)$ of a polygonal path *C* measures the total accumulating turning angle of the counterclockwise tangent as a function of the arclength *s*. The turning angle is measured from some reference point *P0* on *C*, taking the *x*-axis as the reference orientation. Thus, $\Theta_C(0)$ is the angle *v* made by the *x*-axis and the tangent of *C* in *P0*. In order to compare the turning functions of two polygonal shapes, the polygonal length is first scaled so that the total perimeter length is 1. Under this assumption, the distance between two polygons C_1 and C_2 is defined as the L_2-distance between their turning functions $\Theta_{C1}(s)$ and $\Theta_{C_2}(s)$:

$$d_2(C_1, C_2) = \left\| \Theta_{C_1} - \Theta_{C_2} \right\|_2 = \left(\int_0^1 \left(\Theta_{C_1}(s) - \Theta_{C_2}(s) \right)^2 ds \right)^{\frac{1}{2}}, \quad (2)$$

where $\|.\|_2$ is the L_2-norm. The so-defined distance is sensitive to rotation of the polygon C_1 with respect to C_2 and the location of the reference point *P0* on C_1 or C_2. It should be expected, that the polygonal shape of C_i^t and its corresponding polygonal shape in ZC^{t+1} have the same orientation. For this reason the sensitivity of d_2 with respect to rotation will not influence the matching results. To avoid the effects resulting from a bad choice of reference points, it is necessary to set one of the two reference points dependent on the position of the second one. For this aim we distinguish two different cases in our approach: if C_i^t represents a close contour, the reference point *P0* is set arbitrarily on the polygon approximating C_i^t, and the reference point on the candidate polygonal line is chosen as the nearest point to *P0*. Otherwise (when C_i^t is open), *P0* is set in one of the two end-points of C_i^t and the reference point on the candidate polygonal line is chosen as the nearest end-point to *P0*.

To improve the reliability of the similarity measure we not only compare the shape of the two contours but additionally we compare the shape of the one-dimensional discrete function generated by the image intensity when we traverse the whole contour. To avoid effects of noise, we consider the smoothed image $I_G(x,y)$ obtained by convoluting the original image with a Gaussian kernel. Starting at *P0* and traversing C_i^t completely from the start-point *P0* to its end-point *Pl*, the discrete function $I_C : [0, l] \to \Re$ is defined as $I_C(i) = I_G(Pi)$. Note that *l* is the length (in pixels) of the contour, moreover *Pl* lies in the 8-neigborhood of *P0* when C_i^t is closed. The graph of I_C in the $[0, l] \times \Re$-plane defines a contour, which is not necessarily connected in the 8-neighborhood sense. Let us denote this contour as IC_i^t. Due to the noise reduction performed on the image I_G and the fact that abrupt intensity variations along C_i^t cannot happen, it is expected that IC_i^t still remain smooth. Obviously, IC_i^t can be approximated by a polygonal line and its contour length can also be scaled to a total perimeter length of 1. Under this considerations we define the similarity measure between two relevant edges C_i^t and C_j^{t+1} extracted from the image at times *t* and *t+1,* respectively, as:

$$\Phi(C_i^t, C_j^{t+1}) = d_2(C_i^t, C_j^{t+1}) + \alpha . d_2(IC_i^t, IC_j^{t+1}), \quad (3)$$

with $\alpha \in [0,1]$ a weight coefficient. Note that Φ is a well-defined metric between contours.

For a given contour $C_i^t \in ZC^t$ the corresponding contour in ZC^{t+1} is selected as those contour C_j^{t+1} for which the following relation is satisfied:

$$\gamma = \Phi(C_i^t, C_j^{t+1}) = \min_{k \in [0,m]} \Phi(C_i^t, C_k^{t+1}), \text{ and } \gamma \le T,$$

where *T* is a predefined threshold. If $\gamma > T$, then contour C_i^t is declared as unmatchable.

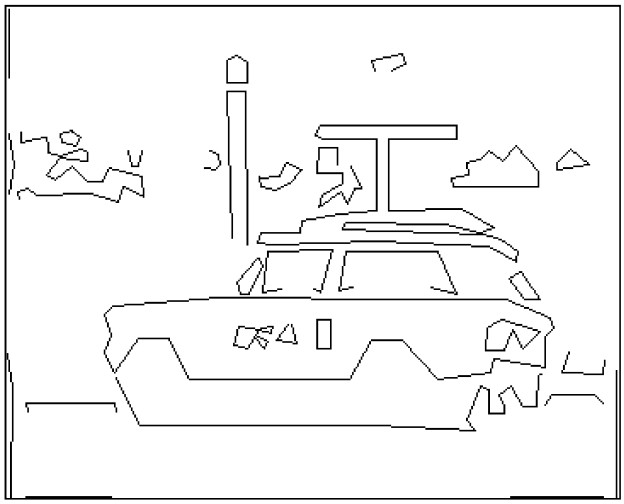

Fig. 3. Polygonal approximation of the edges shown in Fig. 2.

Once correspondences between edges in ZC^{t+1} and ZC^t have been estimated, object masks are extracted using the motion of

pixels along the matched edges. In this last processing step a segment mask is obtained as the connected image area with minimal perimeter containing all relevant edges that undergo the same or similar motion. An especial case should be considered, when the contour of a area extracted according to this condition intersects a unmatchable edge. Let be R such a image region, $C_k^t \in ZC^t$ be the unmatchable edge intersecting the contour of R and S be the set all edges totally contained in R. If the majority of C_k^t lies inside of R, then C_k^t is added to the set S and new image region is extracted by applying the same condition stated above to the set of edges $S \cup C_k^t$. If the majority of C_k^t lies outside of R, then C_k^t is removed from the set ZC^t of relevant edges in the image.

Fig. 4: Extracted foreground segment by grouping and linking edges undergoing the same motion.

5. SIMULATIONS RESULTS

Several experiments have been performed to examine how the presented methods work with real data. In this section we report on the results obtained for the sequence CAR. The image resolution in this sequence is 720x576 pixels. The aim in this experiment is to separate the car moving on the street from the background. For the anisotropic diffusion procedure the discretization of the time interval has been set as $t_j = t_{j-1} + 1$ *for* $j = 1,\ldots,20$ *and* $t_0 = 0$. In the image at the top of Fig. 1 the original 250th frame of the sequence is shown. This image corresponds to the initial scale-parameter t_0. The image shown at the bottom has been obtained for the time instance *t=20*. Fig. 2 shows zero-crossings extracted from the image at the bottom of Fig. 1. Edges were completed using zero-crossings extracted for all diffused images at time t_j for $j = 2,\ldots,20$. The polygonal approximation shown in Fig. 3 has been obtained by applying the technique described at the end of section 3. Here, the width of the δ-band has been set to 5 pixels. Fig. 4 shows the extracted foreground object after grouping and merging relevant edges undergoing the same motion. Correspondences have been estimated between the frames 250th and 252nd (skipping one frame). In this representation intensity values inside of the extracted mask for the foreground object are shown, while the background has been faded out.

6. CONCLUSIONS

A method for motion segmentation has been introduced. The presented research addresses the problem of extract accurate masks of physical objects in moving sequences by using the dynamic of the scene. This goal is achieved by combining edge detection in scale-space and matching of relevant object contours in two different images. To solve the later task a suitable metric for quantization similarity between two edges extracted from different images is introduced. Finally, the performance of the methods presented have been evaluated by processing natural sequences. The extracted object masks approximate quit good the shape of physical objects moving in the scene.

ACKNOWLEDGEMENT

The research leading to this article was supported by the Virtual Centre of Excellence in Digital Broadcasting and Multimedia Technology Ltd., U.K.

REFERENCES

[1] G. Borshukov, G. Bozdagi, Y. Altunbasak and M. Tekalp, "Motion Segmentation by Multistage Affine Classification", *IEEE Transaction on Image Processing*, vol. 6, no. 11, 1997, pp. 1591-1594.

[2] M. Chang, M. Tekalp and, I. Sezan "Simultaneous Motion Estimation and Segmentation", *IEEE Transaction on Image Processing*, vol. 6, no. 9, pp. 1326-1333, 1997.

[3] A. Ibenthal, S. Siggelkow, R. R. Grigat, "Image sequence segmentation for object-oriented coding", *Proc. On European Symposium on Advanced Imaging and Network Technologies*, SPIE vol. 2952, Berlin, Germany, 1996, pp. 2-11.

[4] E. Izquierdo and S. Kruse, " Image Analysis for 3D Modeling, Rendering, and Virtual View Generation", *Computer Vision and Image Understanding, Special Issue on Computer Vision Applications for Network-Centric Computing,* vol. 71, no. 2, 1998, pp. 231-253.

[5] A. Kalvin, E. Schonberg, J. T. Schwartz and M. Sharir, "Two dimensional model based boundary matching using footprints", *Int. J. Robotics Res.*, vol. 5, no. 4, 1986, pp. 38-55.

[6] P. Perona and J. Malik, „Scale Space and Edge Detection Using Anisotropic Diffusion", *Proc. IEEE Comput. Soc. Workshop on Comput. Vision*, 1987, pp. 16-22.

[7] J. Wolfson, "On Curve Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-12, no. 5, 1990, pp. 483-489.