Head-Shoulder Moving Object Contour Tracking using Shape Model

Yong-Ren Huang[1], Chung-Ming Kuo[2] and Chaur-Heh Hsieh[2]
[1]Department of Information Engineering Shu-Te University
[2]Department of Information Engineering I-Shou University
[1]E-Mail: e5646@ms15.hinet.net

Abstract

This paper proposes a new approach for tracking the contour of moving object for head-shoulder video sequence using initial shape model. First, we can utilize manual process or some segmentation approach to obtain the initial shape model. Then, we detect the changing ratio of intensity outside the bounding box in background area. And, the OR operation is employed to obtain a combined shape which is composed of initial shape model and mesh-based temporal segmentation result by motion activity. Finally, we propose a new iterative pixel-based erosion algorithm for obtaining exact object shape using local gradient along the boundary of combined shape in a counting window.

Keyword: moving object contour tracking, shape model, mesh-based temporal segmentation, local gradient.

1. Introduction

For object based video coding system, the segmentation technologies of image sequence, which separate the image into object(s) and background, are necessary and desirable. There are two spatial-temporal and temporal-spatial approaches for video segmentation. In spatial-temporal segmentation [1]-[3], the image is separated into many regions. And objects are obtained by merging process with motion estimation of each region. In temporal-spatial segmentation [4], the motion estimation is processed by pel. And the motion vectors are classified into several regions. In [5], the temporal process is done by motion activity. And the patches with higher motion activity are detected. The silhouette is detected using patterns type in spatial process.

There are several methods is proposed for object(s) tracking [6]-[10]. The object edge detection is proposed by Hausdorff distance [10]. In this approach, the video segmentation result must be obtained in the current frame firstly. Then the spatial segmentation is done in the reference frame. The Hausdorff distance is calculated for the estimation of the edge points of regions between these frames. If the spatial segmentation is not exact enough, the result of the tracking will be influenced. And the computation of the spatial segmentation and the distance measurement are heavy.

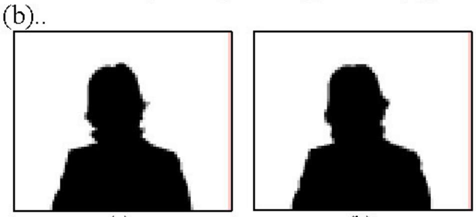
In the object-based video coding system, to acquire the complete object(s) in whole video sequence is required. But the proposed approaches of video segmentation can not segment the object(s) in whole video sequence exactly. So the process of object tracking, after we obtain a complete object model by artificial or segmentation processing, is very important for the object-based video coding system.

In this paper, we propose a new method for moving object contour detection using shape model. We focus our algorithm on head-shoulder video sequences. The new combined shape is obtained from temporal segmentation result and initial shape model by OR operation, which will be discussed in section II part A. the local gradient in a counting window is detected along the real object boundary for the pruning of the combined shape, and an iterative pixel-based erosion algorithm is proposed for obtaining real object shape exactly. The detail algorithm is described in section II part B. And the experimental results will be shown in Section III.

2. Moving Object Contour Tracking using Shape Model

A. Initial shape model and combined shape

The initial shape model for object contour tracking can be obtained by artificial or segmentation approaches. We will utilize these two kinds of initial shape models in our experiment. The artificial shape model is shown in Fig.1(a). And the other one is obtained from temporal-spatial approach [5], shown as Fig.1(b)..



(a) (b)

Fig 1. Shape model for object contour detection.
(a) The artificial.
(b) The temporal-spatial segmentation.

For detecting the moving area of object, we employ temporal segmentation in [5] to obtain the moving regions. The temporal segmentation result may not contain the whole object completely, when the object is with less motion or stationary in a series of frames, shown in Fig.2. So we can not only use the temporal segmentation result for the object contour tracking in reference frame. Now, the shape model and the temporal segmentation result are combined into a new shape image by OR operation, denoted as combined shape. The combined shape contains the object shape in current frame and the moving area of reference frame, denoted in Fig.3.

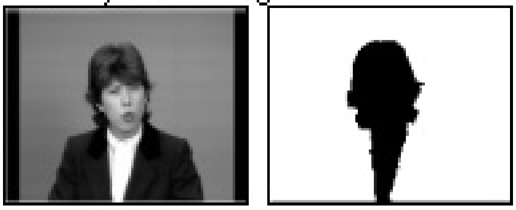


Fig.2 (a) Original image.
(b) The temporal segmentation result with less motion.

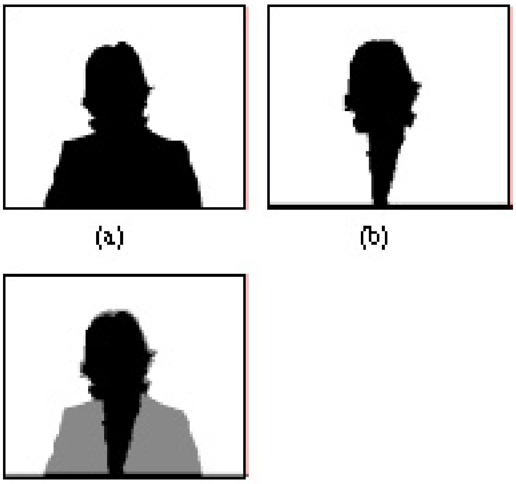


Fig.3 (a) Initial shape model.
(b) Temporal segmentation result.
(c) Combined shape.

B. Pixel-based erosion algorithm for object contour tracking

In the above, the combined shape does not exactly match to the real object boundary yet. A fast pixel-based algorithm proposed for the modification of the combined shape is discussed as follows.

Because the combined shape is composed of shape model and temporal segmentation result, so we must consider the situations of the combination firstly. There are four states between them. These states are containing, covering, overlapping and covered, shown as Fig.4. The combined shape, denoted as S_c, which is composed of initial shape model S_m and temporal segmentation result S_t.

If the combination situation is in containing state, we keep the original shape model as the real object shape in the reference frame while the ratio of the temporal segmentation result area to shape model area is less than 80%.

If the overlapping or covered state occurs, we create a fast pixel-based erosion algorithm for obtaining accurate object shape. First step, the change of luminance in an image sequence may be caused by environment, such as sunlight, flash lamp, fade in, fade out etc. We find the bounding box of real object firstly. Then we calculate the changing ratio of intensity of the current image to the reference outside the bounding box, denoted as L_r.

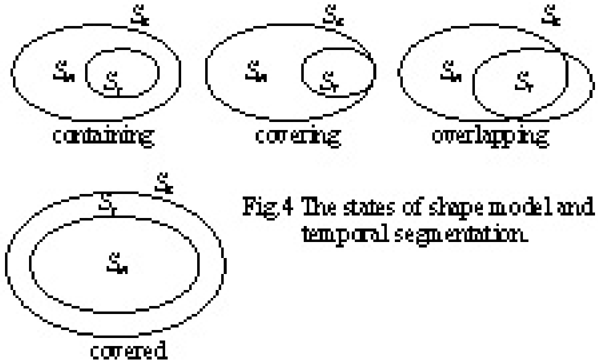


Fig.4 The states of shape model and temporal segmentation.

Second step, let $C=\{c(x,y,t)|0 \le t \le N_f\}$ is the set of points located inside along the edge of the shape at time t, where N_f denotes the number of frames. And $\hat{C}=\{\hat{c}(x,y,t)\ |0 \le t \le N_f\}$ denotes the set of points outside neighboring to the edge point (x,y) of the shape. And let $I(c(x,y,t))$ denote intensity value at location $c(x,y,t)$ in image. And the $\nabla I(c(x,y,t))$ is the average value of gradient within 4-connectivity neighboring points.

In forward object contour tracking, the contour point $c(x,y,t)$ is in the initial shape model, and the point $c(x,y,t+1)$ is in the combined shape which is obtained from the combination of initial shape model and temporal segmentation result. We have to decide whether the point is the contour point of real object in reference frame or not. So, we define a counting window which size is ±8 centered at contour point. If the whole counting window is out of the shape model, the contour point of the combined shape is erased. Otherwise, the average gradient value denoted as $\nabla I^t(x,y)$ is calculated in the following equation.

$$\nabla I(c(x,y,t)) = \frac{1}{Np}\sum_{i=x-8}^{x+8}\sum_{j=y-8}^{y+8}|I(c(i,j,t)) - I(\hat{c}(i,j,t))| \quad (1)$$

where Np is the number of contour points located in $\{c(i,j,t)\,|\,x\text{-}8 \le i \le x+8, y\text{-}8 \le j \le y+8\}$. And then the gradient value $\nabla I(c(x,y,t+1))$ is calculated as follows.

$$\nabla I(c(x,y,t+1)) = \frac{Lr}{Nb}\sum|I(c(x,y,t+1)) - I(\hat{c}(x,y,t+1))| \quad (2)$$

where Nb is the number of neighboring points to $c(x,y,t+1)$. If $\nabla I^t(x,y) > \nabla I^{t+1}(x,y)$, it means that the contour point $c(x,y,t+1)$ of the combined shape is belonged to background. So, we erase this contour point.

If $c(x,y,t+1)=c(x,y,t)$, the covering state occurs. We still can not determine that the point is edge point of real object or not, because there may be a slight motion in the area. The Mean Square Error values, denoted as Mi and Mo expressed in Eq.(3)(4), are calculated for determining the similarity of these points firstly.

$$Mi(x,y) = (I(c(x,y,t)) - I(c(x,y,t+1)) \cdot Lr)^2 \quad (3)$$

$$Mo(x,y) = (I(\hat{c}(x,y,t)) - I(c(x,y,t+1)) \cdot Lr)^2 \quad (4)$$

If Mo>Mi, we keep this point as contour. Otherwise, The gradient $\nabla I^t(x,y)$ and $\nabla I^{t}(x,y)$ are calculated.

$$\nabla I^t(x,y) = \frac{1}{Nc}\sum|I(c(x,y,t)) - I(\hat{c}(x,y,t))| \quad (5)$$

$$\nabla I^t(x,y) = \frac{Lr}{Nr}\sum|I(c(x,y,t+1)) - I(\hat{c}(x,y,t+1))| \quad (6)$$

where Nc is the number of points of $\hat{c}(x,y,t)$ and Nr is the number of points of $\varepsilon(x, y, t+1)$. If $\nabla I'(x,y) > \nabla I(x,y)$, the contour point is erased. And the second step is executed iteratively, until the combined shape does not change.

In successive simulation of head-shoulder object contour tracking, we update the initial shape model using the obtained shape for object contour tracking in following frame.

3. Experimental Results

The test video sequences Claire and Trevor in size of 352x240 and 256x256 respectively. We test the tracking algorithm on the computer which CPU is in the speed of 1GHz. And the Table I is the average spending time in the contour detection process.

In our experimental results, we consider the initial shape model by artificial for the contour detection, shown in Fig.5. And the results are accurate on the real object boundary, denoted in Fig.6.

4. Conclusions

This paper presents a new method for moving object contour tracking using initial shape model. We consider the change of background intensity firstly. And we acquire a combined shape from mesh-based temporal segmentation and initial shape model by OR operation. We consider four states of the composition. Finally, we create a pixel-based erosion algorithm for obtaining accurate object contour. We calculate local gradient in a counting window along the real object boundary to determine the points erasing or not. And iterative processing the erosion algorithm until there is no change of shape. This algorithm is simple to implement and the tracking process is fast.

References

[1] Demin Wang, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking", IEEE Trans. on Circuits and Systems for Video Tech., Vol 8, No. 5, pp. 539-546, Sept. 1998.

[2] Jae Gark Choi, Si-Woong Lee and Seong-Dae Kim, Member, "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure", IEEE Trans. on Circuits and Systems for Video Tech., Vol 7, No. 2, pp. 279-286, Apr. 1997.

[3] Ioannis Patras, E.A. Hendriks and R.L. Lagendijk, "Video Segmentation by MAP Labeling of Watershed Segments", IEEE Trans. on Pattern Anal. and Machine Intell., Vol. 23, No. 3, pp. 326-332, Mar. 2001.

[4] Michael M. Chang, A. Murat Tekalp, and M. Ibrahim Sezan, "Simultaneous Motion Estimation and Segmentation", IEEE Trans. On Image Processing, vol. 6, NO. 9, pp. 930 -938, Sept. 1997.

[5] Chung-Ming Kuo, Chaur-Heh Hsieh and Yong-Ren Huang, "Automatic Extraction of Moving Objects for Head-Shoulder Video Sequence," Journal of Visual Communication and Image Representation, Vol 16, No. 1, Feb. 2005.

[6] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar, "Object Tracking with Bayesian Estimation of Dynamic Layer Representations ", IEEE Trans. on Pattern Anal. and Machine Intell., Vol. 24, No. 1, pp. 75-89, Jan. 2002.

[7] Johnson I Agbinya and David Rees, "Multi-Object Tracking in Video", Real-Time Imaging Issue 5 , pp. 295-304, Oct. 1999.

[8] Jianping Fan, Xingquan Zhu, and Lide Wu , "Automatic Model-Based Semantic Object Extraction Algorithm", IEEE Trans. on Circuits and Systems for Video Tech., vol. 11, NO. 10, pp. 1073-1084, Oct. 2001.

[9] Thomas Meier and King N. Ngan, "Automatic Segmentation of Moving Objects for Video Object Plane Generation", IEEE Trans. on Circuits And Systems for Video Tech., Vol 8, No. 5, pp.525-538, Sept. 1998.

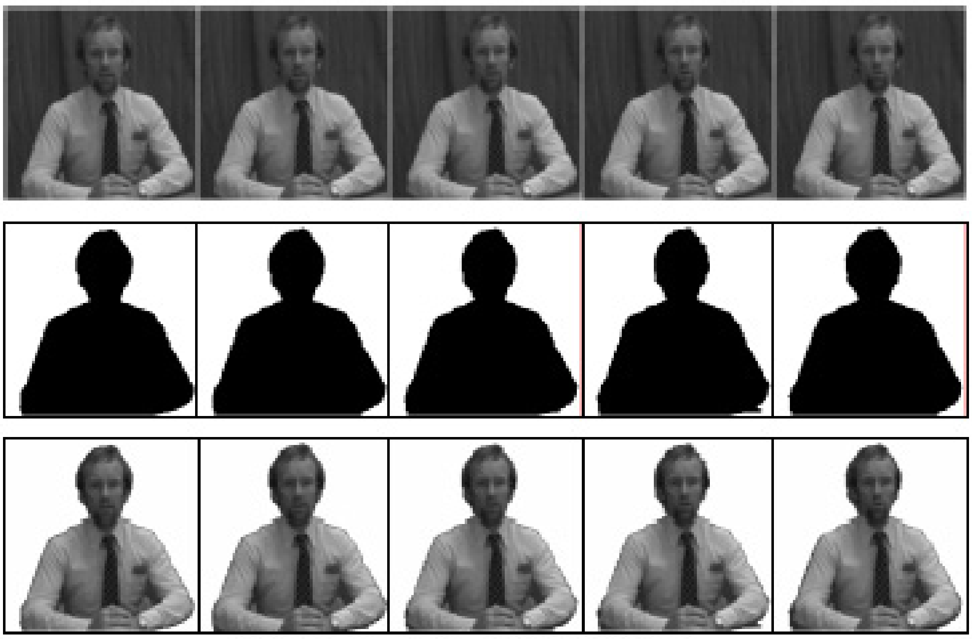
[10] D. P. Huttenlocher, G. Klanderman, andW. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Trans. Pattern Anal. and Machine Intell., Vol 15, pp. 850-863, 1993.

Table I. The average spending time of the contour detection.

Video sequence	Temporal segmentation per frame	Contour tracking per frame	Total time
Claire	391ms	281ms	672ms
Trevor	321ms	521ms	842ms

Fig. 5 (a) Real image and the artificial shape model of the 128 th Claire sequence.
(b) Real image and the artificial shape model of the 34 th Trevor sequence.

(a) The results of 128'th ~ 132'th Claire sequence.



(b) The results of 34'th~38'th Trevor sequence.

Fig.6 The results of object contour tracking.