ACCURATE MOVING OBJECT SEGMENTATION BY A HIERARCHICAL REGION LABELING APPROACH

Wei Zeng[1] *and Wen Gao*[1, 2]

[1]Department of Computer Science and Technology, Harbin Institute of Technology, China;
[2]Graduate School of Chinese Academy of Sciences, China
Email: {wzeng, wgao }@jdl.ac.cn

ABSTRACT

This paper proposes a new algorithm to segment moving objects from color sequences accurately. The segmentation procedure is treated as a Markovian labeling process and formulated by a hierarchical Markov random field (MRF) model. At beginning, the original frame is partitioned to homogeneous regions with different granularity by the rapid watershed algorithm. Then foreground is detected as outliers of the estimated background motion in the initial motion classification stage. After that, motion vector is estimated for each foreground region and is validated by an elaborate occlusion detection scheme. The initial object mask is segmented by the MRF model on the larger-scale spatial partition and is refined by the other MRF model in the small-scale partition. The hierarchical MRF models provide the fine object boundary. The proposed method is evaluated on several real-world image sequences and the experimental results shows remarkable performance.

1. INTRODUCTION

Extracting meaningful entities from visual data provides object-based representation for video content. Such semantic representation enriches the means of accessing and manipulating video content conveniently. A variety of applications, range from video compression to video retrieval, from video surveillance to video editing, and from scene analysis to pattern recognition, benefit from the object-based representation. Industry standards, such as the MPEG-4 and MPEG-7, need automatic segmentation tools to support the content-based or object-based video coding and description framework.

Up to now, segmentation of objects in image sequences is still an open problem. A number of techniques have been developed for object extraction in the past years [1-12]. According to whether human aids the segmentation process or not, these techniques are classified into two categories: supervised and unsupervised segmentation approaches. The semiautomatic object segmentation approaches, referred as supervised segmentation, introduce human assistance in the initial object extraction for the key frame [1,2]. On the contrary, the automatic object segmentation approaches, referred as unsupervised segmentation, extract object without any interaction. The change-based approach segments object by detection the pixel or region that disobeys the distribution of background on frame differences signal [3]. Background registration technique sets up the background among multiple frames. The foreground object is segmented by comparison the current frame with the pre-computed background model [4]. Through computing motion vector field, object can be directly segmented by similarity of motion vectors [5]. To obtain an accurate object boundary, color is adopted to refine the object mask [6,7]. To incorporate the object and image region's boundary, spatial partition is utilized to guide the location of segmented object's boundary [8-10]. The interesting work that segments moving object from the compressed bit stream is occurred in real-time applications recently [11,12].

In this paper, we propose a novel hierarchical Markov random field (MRF) model to segment moving objects from image sequences accurately. One MRF model defined by color and motion is used to initialize the object mask on a large-scale spatial partition of the frame. Further, the object boundary is refined by the other MRF model on the small-scale spatial partition. The hierarchical segmentation scheme is consistent with human vision system that segments object in a multi-scale way. Experimental results show remarkable performance in several real sequences.

2. SPATIAL PARTITION

The gradient watershed algorithm is implemented to segment frame into connected spatial regions [13]. This approach has several advantages. First, the boundaries of gradient watersheds correspond quite closely to the edges of the original image. Second, object edge obtained by calculating gradient watershed boundaries are always guaranteed to be connected and closed. While the watershed algorithm produces very accurate result, it has inherent drawback; it is sensitive to gradient noise, and results in over-segmentation.

Two scale's spatial partitions are obtained by the watershed algorithm. The small-scale spatial partition is performed by segmentation the original frame, while the large one is obtained from segmentation the smoothed frame filtered by a Gaussian filter.

3. MOTION CLASSIFICATION

The task of motion classification is to detect foreground regions that have different motion from background. Supposing the background motion is the dominant motion, the foreground is the outlier of the estimated background motion model. The dominant motion is calculated by the robust regression for parametric motion estimation technique [14]. Six-parameters affine model is estimated by a coarse-to-fine gradient-based method using 3-level multi-resolution pyramid. The algorithm is implemented using a robust M-estimation technique to make the computation insensitive to outliers that result from multiple motions.

Motion classification detects foreground regions that violate the background motion model. The weight to each pixel is induced from the M-estimator and represents the likelihood of background motion. Region likelihood is approximated by the average weight within the region. However, the problem is arisen. The low texture parts within a moving region have small weight because of motion uncertainty. The moving region weight will be decreased and may be misclassified as background. Therefore, we developed a two-thresholds detection technique to discover these moving regions robustly. Moving regions are detected by a threshold on the small-scale spatial partition at first step. All pixels within the moving region are marked as moving pixels. After that, the region detection is performed on the large-scale partition. If the number of moving pixel within a region is above the predefined threshold, the region is classified as foreground. By this means, moving regions can be successfully detected, while some noise regions are restrained.

4. MOTION VALIDATION

The occlusion region will be detected as foreground because there is no matched region in the successive frame. This poses a serious problem, since covered background would be classified as foreground. To solve this problem, foreground region motion estimation and validation process is introduced. As the occlusion usually occurred around the object boundary, motion estimation and validation are simply performed on the boundary region.

In theory, occlusion has no matched pixel or region in the successive frame. Imposing a motion vector to an occlusion region, the motion compensated differences (MCDs) will be high. We call the occlusion region as unmatched occlusion. Otherwise, we call the occlusion region as wrong matched occlusion. This occurs when some background is matched with the occlusion region by their similar texture. In unmatched occlusion region, such occlusion regions can be detected by examining the pixel MCDs. Pixel MCD is tested by the hypothesis test based on the statistics of motion matching error. If the number of unmatched pixel is above the predefined threshold, the region is classified as an unmatched occlusion region. The region motion vector is estimated by the robust regression technique for the translation motion model.

However, the wrong matched occlusion is often occurred, because the occlusion is frequently matched with similar background incorrectly. Motivated by the fact that overlap is usually happened between the occlusion region and background in the reference frame after motion projection, a detection scheme detecting overlap followed by the MCD comparison, is developed to recognize the wrong matched occlusion. First, the candidate occlusion region is found by checking whether the motion projected target region is overlapped or not by a motion projected background. Second, MCDs in occlusion region and background are compared pixel by pixel. If the number of pixel's MCD is smaller in the candidate region than background, the region is classified as a wrong matched occlusion region.

5. THE HIERACHICAL MRF MODEL

Object segmentation can be treated as a Markovian labeling process. The labeling process is guided by the maximum a-posteriori (MAP) criterion, and worked as minimization of the following energy function

$$\hat{f} = \arg\min_{f} U(f|d),$$

where $U(f|d)$ is the posterior energy function. f is the label set, $f = \{F, B\}$; d is the set of observations on each grid. In object segmentation, only the foreground label F and the background label B are considered. As the frame has been partitioned into spatial regions, the labeling is performed on regions in this paper. This leads to a region-based segmentation approach.

Moving object segmentation is accomplished by the hierarchical MRF structure. In our implementation, two types of MRF models are defined to capture different visual features on variant scale's partitions of the frame. One MRF model labels the object region by employing motion and color feature on the large-scale partition. This MRF model provides initial segmentation based on the hybrid criteria of motion and color. The other MRF model defined on the small-scale partition refines the object mask boundary by color and smoothness characters. The connection between the two MRF models is a transition-probability-like term defined in the second MRF model. The advantages of the proposed approach are:

- large-scale spatial partition is more consistent with the adjacent relation of spatial parts for real objects;
- the motion vector is robust for large-scale region, and it is very helpful to deficient motion classification.

Assuming the region has been marked with motion label *M* in the motion classification stage, the MRF model defined on the large-scale partition is formulated as

$$U(f|d) = \sum_i V_i^M(f,d) + \sum_{(i,j)\in C_i} V_{i,j}^C(f,d), \qquad (1)$$

where C_i is the clique of location *i*. $V_i^M(f,d)$ is the data-driven energy term for motion, and written as

$$V_i^M(f,d) = \begin{cases} -\alpha_1, & \text{if} \quad f_i = M_i \\ +\alpha_2, & \text{otherwise} \end{cases}.$$

$f_i = M_i$ means that the region label *f* is coincided with the motion label *M*. $V_{i,j}^C(f,d)$ is the spatial continuity term, which describes the color similarity of the adjacent regions around R_i and is formulated as

$$V_{i,j}^C = \begin{cases} -\beta_1 \cdot C(R_i,R_j)\cdot G(R_i,R_j)B_{i,j}, & f_i = f_j \\ +\beta_2 \cdot C(R_i,R_j)\cdot G(R_i,R_j)B_{i,j}, & f_i \neq f_j \end{cases},$$

where $C(R_i,R_j)$ is the color distance function between the adjacent region R_i and R_j. In our implementation, we use the Euclidean distance in HSV color space. $G(R_i,R_j)$ is the gradient distance function that is the average of absolute intensity difference along the common boundary. $B_{i,j}$ is the common boundary length between the two regions.

The MRF model defined on the small-scale partition is

$$U(f|d) = \sum_i V_i^F(f,d) + \sum_{(i,j)\in C_i} V_{i,j}^C(f,d) + \sum_{(i,j)\in C_i} V_{i,j}^S(f,d), \qquad (2)$$

where $V_i^F(f,d)$ is the energy term induced from the segmentation result from the MRF model in E.q.(1). This term defines the occupation ratio between the small-scale partitioned region and its father region on the large-scale partition. $V_i^F(f,d)$ is defined as

$$V_i^F(f,d) = \lambda \frac{S_k}{S_j},$$

where S_k and S_j are label counts for the small-scale partitioned region and its father in the large-scale partition. The highest value of the occupation ratio means that the two regions have the same labels as many. $V_{i,j}^S(f,d)$ is the spatial smooth term, which is defined as

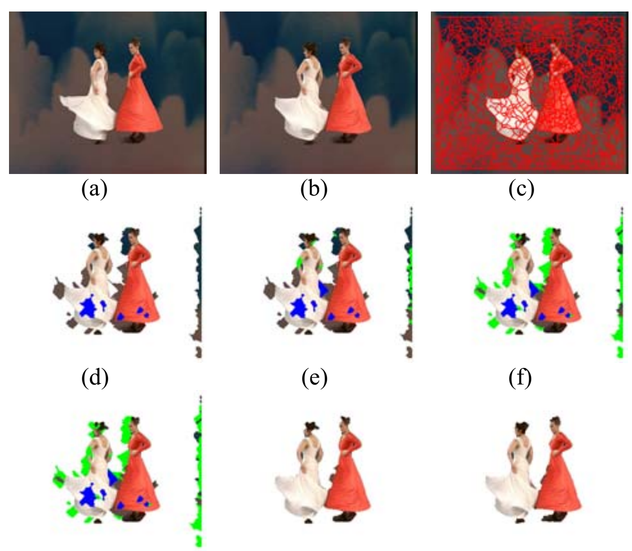
$$V_{i,j}^S = \begin{cases} -\gamma & f_i = f_j \\ +\gamma & f_i \neq f_j \end{cases}.$$

The constants $\alpha,\beta,\gamma,\lambda$ determine the relative contribution of the terms to the energy function. It should be noticed that the spatial smooth term only occurred in the MRF model on the small-scale partition in E.q.(2). This is because that the MRF model on the small-scale partition is used to get the accurate boundary. This term can guarantee to obtain the smooth boundary. The minimization of the energy function is performed by the iterative deterministic relaxation algorithm known as highest confidence first (HCF).

6. EXPERIMENTAL RESULTS

The performance of the proposed algorithm was evaluated on several test sequences in the MPEG4 test set. The sequence *Stefan*, *Dancer* and *Hall*were used. The sequence *Stefan* exhibits a moving camera and the sportsman is presenting a non-rigid motion. The sequence *Dancer* is a synthesized sequence that two dancer are spinning above a panning sky. The sequence *Hall* is typical video conference scene which exhibits a slow and smooth motion over a stationary background.

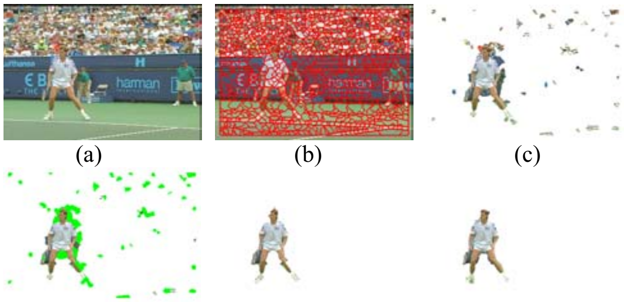
Fig.1 illustrates the segmentation process for the *Dancer* sequence. Two successive frames are shown in Fig.1 (a) and Fig.1 (b). The spatial partition at large scale is given in Fig.1 (c). Based on this partition, the motion classification result is obtained and depicted in Fig.1 (d), where the detected background is marked with white color. Fig.1 (e) shows the unmatched occlusion regions that marked with green color. Fig.1 (f) demonstrates wrong matched occlusion regions. Fig.1 (g) displays the detected occlusion regions. The MRF classification on the large-scale partition provides a good initialization for object segmentation (Fig.1 (h)), and the refined moving object mask is shown in Fig.1 (i). In the ultimate segmentation mask, object boundary is more accurate.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

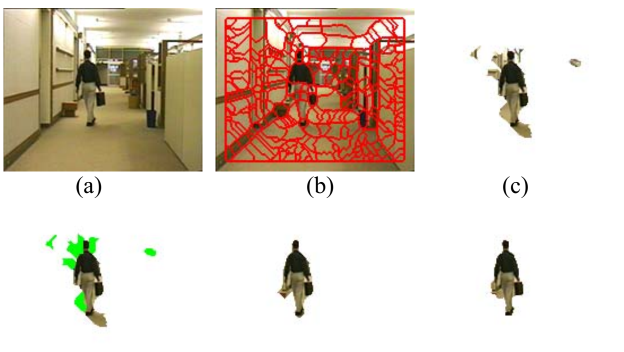
Figure 1. Segmentation process for the 141th frame of *Dancer*.

Satisfactory result was achieved in *Stefan* sequence as shown in Fig.2. Details are found by the refining process, especially at right shoe and left calf of the sportsman. The result for *Hall,* shown in Fig.3, is somewhat less accurate. The result is mainly due to the fact that the foreground motion is very slow.



(a) (b) (c) (d) (e) (f)

Figure 2. Segmentation of the 48th of *Stefan*. (a)the original 48th frame, (b)the large scale spatial partition, (c)the result of motion classification, (d)occlusion regions, (e)MRF classification on large-scale partition, (f)MRF classification on small-scale partition.



(a) (b) (c) (d) (e) (f)

Figure 3: Segmentation of the 57th of *Hall*. (a)the original 57th frame, (b)the large scale spatial partition, (c)the result of motion classification, (d)occlusion regions, (e)MRF classification on the large-scale partition, (f)MRF classification on the small-scale partition.

7. CONCLUSION

This paper proposes a novel algorithm to extract moving objects from image sequences. The hierarchical MRF model is used to segment object and refine boundary on different scale's spatial partitions. The performance of the proposed algorithm is evaluated on several real sequences and achieves the remarkable results.

8. ACKNOWLEDGEMENT

Thank for Dr. Xilin Chen's valuable discussion. This work has been supported by National Hi-Tech Development Programs of China under grant No. 2003AA142140.

9. REFERENCES

[1] C. Gu, and M.C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," *IEEE Trans. on Circuits Syst. Video Technol.*, vol.8, no.5, pp.572-584, Sept. 1998.

[2] S.Sun, D.R.Haynor, and Y.Kim, "Semiautomatic video object segmentation using vsnakes," *IEEE Trans. on Circuits Syst. Video Technol.*, vol.13, no.1, pp.75-82, Jan. 2003.

[3] A.Neri, S.Colonnese, G.Russo, P.Talone,"Automatic moving object and background separation," *Signal Processing*, vol.66, pp. 219-232, 1998.

[4] S.Y. Chien, S. Yih, and L.G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. on Circuits Syst. Video Technol.*, vol.12, no.7, pp.577-586, Jul. 2002.

[5] A.R.Mansouri and J.Konrad,"Multiple motion segmentation with level sets," *IEEE Trans. on Image Process*., vol.12, pp.201-220, Feb.2003.

[6] A.Cavallaro, T.Ebrahimi,"Accurate video object segmentation through change detection", *Proc. of IEEE Intl. Conf. on Multimedia and Expo*, pp.445-448, Aug. 2002.

[7] M.Piek, R.Braspenning and C.Varekamp,"Unsupervised motion-based object segmentation refined by colour," *Proc. of SPIE on Visual Communications and Image Processing*, Lugano, Switzerland, vol.5105, Jul. 2003.

[8] D.G. Perez, C.Gu and M.T Sun, "Semantic video object extraction using four-band watershed and partition lattice operators," *IEEE Trans. on Circuits Syst. Video Technol.*, vol.11, no.7, pp.788-801, Jul. 2002.

[9] Y.Tsaig and A.Averbuch,"Automatic segmentation of moving objects in video sequences: a region labeling approach," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 12, no.7, pp.597-612, Jul.2002.

[10] C. Marc, P. Stephane and N. Henri,"Segmentation of non-rigid video objects using long term temporal consistency," *Proc. of IEEE Intl. Conf. on Image Processing*, Rochester, New York, Spain, Sept. 2002.

[11] W. Zeng and W. Gao, "Automatic object extraction in mpeg video", Proc. of IEEE Intl. Symposium on Circuits and Systems, Bangkok, Thailand, vol.II, pp. 524-527, May 2003.

[12] Z. Liang, P.Wang and Z.Tan, "Moving object detection from mpeg bit stream," *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Hong Kong, China, vol.III, pp.701-704, Apr. 2003.

[13] J.M.Gauch, "Image segmentation and analysis via multiscale gradient watershed hierarchiers," *IEEE Trans. on Circuits Syst. Video Technol*.,vol.l8, no.1, pp.69-79, Jan.1999.

[14] M.J. Black, and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol.63, pp. 75-104, Jan.1996.