AN ALGORITHM FOR SEGMENTING MOVING VEHICLES

Su Zhang[1,2], *Hanfeng Chen*[1], *Zheru Chi*[2], *Pengfei Shi*[1]

[1] School of Electronics & Information Technology
Shanghai Jiao Tong University
Shanghai, P.R. China

[2] Center for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

An algorithm for region-based moving object segmentation is presented in this paper. The gray-scale image segmentation based on the Mean Shift Algorithm (MSA) is performed first to segment each frame of a sequence into connective homogeneous regions. A method applying spatio-temporal continuity constrains of the motion vector image is then carried out to detect moving pixels robustly. Finally, each homogeneous region is labeled as either a moving-object region or a non-moving-object region according to the number of moving pixels it contains. Experimental results show that our algorithm is effective and robust in segmenting moving vehicle from noisy scenes.

1. INTRODUCTION

Object-based video representation, in terms of semantically meaningful moving 2-D layers, has recently become popular for many applications, including monitor system, object recognition, video compression and content-based access. The precondition of object-based video representation is video objects segmentation. This paper is focused on moving vehicles segmentation in some monitor systems. In these systems, the objects are moving while the background is static.

Satisfied segmentation of moving objects from a static scene is often difficult because of noise. Another challenge is known as blank wall problem [1]. The motion vectors of inner pixels of a homogeneous region with uniform color are difficult to be decided.

In recent years a number of different approaches have been proposed for moving objects segmentation. An segmentation algorithm based on an object-background probability estimation is presented in [2]. The algorithm assumes that moving objects and background are simply connected and no less than one point is known beforehand to belong to the background.

Region-based moving objects segmentation [3][4] is a convenient and robust segmentation algorithm where each object is assumed to be connected with various homogeneous regions. Image I is supposed to be composed of regions $\{R_1, R_2, ..., R_n\}$. The problem of region-based moving objects segmentation is to get the subset $\{R_1, R_2, ..., R_m\}$ of $\{R_1, R_2, ..., R_n\}$, where $\{R_1, R_2, ..., R_m\}$ is the collection of all moving objects regions. In the region-based segmentation each frame is segmented into various homogeneous regions firstly. This step is a low-level computer vision task which intends to contribute to the later semantic-level segmentation. Then, pixels in each image are classified into moving pixels and static pixels. Finally, moving objects are segmented by grouping those homogeneous regions including a certain ratio of moving pixels. Figure 1 illustrates the framework of region-based moving objects segmentation.

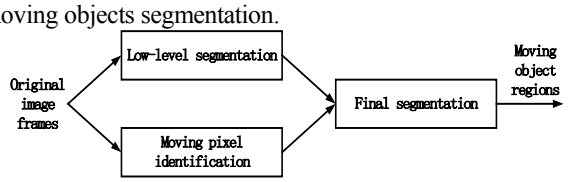


Figure 1: Block diagram of region-based moving object segmentation

The efficiency of region-based segmentation relies on the low-level segmentation and moving pixels detecting methods. The usual solution of low-level segmentation is to segment each original image into grey-homogeneous regions by region growing [3][4]. Pixels of each region are of the same grey value and each region is spatially connective. The grey level of original image may be quantified evenly before region growing. However, the resolution of the segmentation may be too high because of rich texture and noise, which results in too many small regions. It will decrease the accuracy of later high-level segmentation, since the final segmentation is performed by grouping these regions. Mean shift algorithm (MSA) is introduced in this paper to perform the homogeneous regions segmentation. MSA has been exploited in some color image segmentation cases [5]. This paper will show that MSA can contribute to moving objects segmentation in video sequence.

Moving pixels are those pixels belonging to moving objects regions. The thresholded image difference [6] of two neighboring frames is in common use to detect moving pixels. Unfortunately, difference image is sensitive to noise. Block-based motion estimation is used in this paper to detect moving pixels. Motion vectors of pixels is estimated by block-matching. Then, spatio-temporal continuity constrains of motion is imposed to refine these motion vectors more. While block-matching based on the translatory block model is simple, it deals poorly with rotations and deformations of blocks from frame to frame. Fortunately, precise motion vectors are not required strictly since our goal is only to clarify pixels into moving pixels or background pixels.

The paper is structured as follows. Section II introduces the mean shift algorithm (MSA) and grey image segmentation based on MSA. Section III proposes a new method to identify moving pixels, while the final segmentation algorithm is presented in section IV. The experimental results are shown in Section V. Finally, section VI draws the conclusions.

2. IMAGE SEGMENTATION BASED ON MSA

2.1. Mean Shift Algorithm (MSA)

Feature space analysis is a widely used tool for solving low-level image understanding tasks. Significant features in the image correspond to high density regions in the feature space and the highest density regions correspond to cluster centers. The mean shift algorithm was proposed as a method for cluster analysis [7]. The MSA is an iteration process in finding the cluster center. The basic idea of the MSA is to move a shift window in the gradient direction of the feature space, starting at a randomly selected point. The convergence point of the shift window center is a cluster center.

Let S be a finite set embedded in the n-dimensional Euclidean space X and let K be a flat kernel that is the characteristic function of the λ-ball in X,

$$K(x) = \begin{cases} 1 & \text{if } \|x\| \le \lambda \\ 0 & \text{if } \|x\| > \lambda \end{cases}. \quad (1)$$

Then, the sample mean with kernel K at $x \in X$ is defined as[9]

$$m(x) = \frac{\sum_{s \in S} K(s-x)w(s)s}{\sum_{s \in S} K(s-x)w(s)} \quad (2)$$

where $w : S \to (0, \infty)$ is a weight function. Let $d(x)$ the difference of $m(x)$ and x,

$$d(x) = m(x) - x \quad (3)$$

$d(x)$ is called mean shift. For each $x \in X$ there is a sequence $\{x, m(x), m(m(x)), \ldots\}$ which is called the trajectory of x. The evolution of x in the form of iterations $x \leftarrow m(x)$ with $x = m(x)$ is called a mean shift algorithm. The iteration will converge to a fixed point $msa(x)$,

$$msa(x) = \lim_{n \to \infty} m^n(x) \quad (4)$$

where $m^n(x) = m(m^{n-1}(x))$. $msa(x)$ is considered a cluster center of the space X. $msa(X) = \{ msa(x), x \in X \}$ is the collection of all the cluster centers. It should be emphasized that $msa(x_1)$ may be equal to $msa(x_2)$ though $x_1 \neq x_2$.

In fact, the kernel K could be varied, such as Gaussian kernel and Epanechnikov kernel [8]. If K is defined as (1), it provides a shift window of the radius λ and $m(x)$ is the barycenter of the window centered at point x if $w(s)$ is defined as the density function of the feature space. According to (2), the shift window moves in the gradient direction of the function g on X,

$$g(x) = \sum_{s \in S} K(s-x)w(s) \quad (5)$$

The moving distance of each iteration step is $\|d(x)\|$. The convergence of the MSA has been verified in [8].

2.2. Grey-scale Image Segmentation Based on MSA

Low-level image segmentation partitions an image I into homogenous regions $\{R_1, R_2, ..., R_n\}$, where $I = \bigcup_{i=1}^{n} R_i$. The usual solution of this task is to segment original image into gray-homogeneous regions by region growing. Such a simple mechanism utilizes only the spatial neighbor relation of pixels in the image domain while rich statistic characters in the feature space are wasted. This method is consequently sensitive to rich texture and noise and results in over-segmentation. Over-segmentation goes against solving the challenge of blank wall problem, as shown in experimental results.

Dorin [5] proposed color image segmentation algorithm based on MSA. In this paper, the algorithm is introduced to the region-based moving objects segmentation as the low-level gray image segmentation. For a gray image, the gray histogram is adopted as its feature space in this paper.

A gray image could contain 256 different gray levels theoretically. However, only a few main gray levels are magisterial in an actual gray image usually. These main gray levels are cluster centers which corresponding high density regions in the histogram. MSA is performed to get these centers and a palette composed of these main gray levels is created. The palette provides the gray levels allowed when segment the gray image. Gray levels of all the pixels are reallocated based on the palette in the segmentation and minor gray levels are deleted from the image.

The steps of gray image segmentation based on MSA are presented below:

(1) Definition of the segmentation parameters.

Four parameters should be set before segmentation. They are the width of the shift window in MSA λ, the threshold r indicating the end of MSA iteration, the smallest number $NUM1$ of pixels required for a major gray level and the smallest number $NUM2$ of contiguous pixels required for a homogenous region.

(2) Map the image domain into the feature space, histogram.

(3) Initialization of the shift window.

A few, such as 25, pixels are sampled randomly as candidates in the image domain. For each pixel, the mean of its 3×3 neighborhood is computed and mapped into the histogram. The window containing the highest density of histogram is selected as the initial shift window.

(4) Get a major gray level with MSA.

Started at the initial shift window, MSA is performed and convergence is declared when $\|d(x)\| \le r$. The center of the shift window in the histogram is a major gray level. If the number of pixels with this gray level exceeds $NUM1$, this gray level is added to the palette.

(5) Modification of the original image and histogram.

Pixels with the gray levels inside the shift window and their 8-connected neighbors are deleted from the image domain. Then, the histogram should be modified since many pixels have been removed in the image domain.

(6) Steps 3 to 5 are repeated until no gray level is added to the palette longer.

(7) Gray level of each pixel is allocated to the nearest gray level in the palette.

(8) Region growing.

After step 7, only major gray levels are present in the image. Then traditional region growing is performed to segment the image into homogenous regions $\{R_1, R_2, ..., R_n\}$. Finally, each region holding less than *NUM2* pixels is emerged to one of its neighbor regions according to the gray level similarity as the post-processing.

3. DETECTION OF MOVING PIXELS

The usual methods to detect moving pixels in a static background are based on thresholding image difference of each two neighboring frames[6]. Unfortunately, image difference is sensitive to noise. Some statistics methods, such as high order statistics (HOS) [9], were developed to reduce the effect of additive noise. However, these methods are also based on image difference. In this paper, moving pixels are detected initially by block-based motion estimation. Then, the spatio-temporal continuity constrains of motion are considered to refine the moving pixels.

Let I_v[*i,j*] be the motion vector of pixel (i, j). The original detection of moving pixels can be described as

$$\boldsymbol{omp}[\boldsymbol{i},\boldsymbol{j}]=\begin{cases}1 & |\boldsymbol{I}_v[\boldsymbol{i},\boldsymbol{j}]|\neq 0\\ 0 & |\boldsymbol{I}_v[\boldsymbol{i},\boldsymbol{j}]|= 0\end{cases} \tag{6}$$

where *omp* is a matrix of the same width and height with the original image. o*mp*[*i,j*] equals to 1 if pixel (i, j) is judged as a moving pixel and 0 if background pixel.

However, "moving pixels" detected by *omp* may include many background pixels because of noise. Motion vectors caused by noise are usually distributed on some isolated small regions and vary acutely along the time. Thus, spatio-temporal continuity constrains of motion is needed to refine the moving pixels. The spatio-temporal continuity constrains of motion is based on two easy assumptions. Firstly, spatially neighbored pixels are assumed of similar motion vectors. Secondly, the motion vector of an actual point is assumed to change smoothly during a brief time, such as the time going through three frames. Exceptions to the first assumption may occur on the bonder of each object. Also, some pixels may disobey the second assumption if they belong to objects starting moving or to stop. Fortunately, these exceptions are usually minor considering all the moving pixels and can be reduced greatly in later process.

Let I_{hv} and I_{vv} be the horizontal vector image and vertical vector image of the motion vector image I_v, which means I_{hv}[*i,j*] and I_{vv}[*i,j*] are elements $\hat{d}_1$ and $\hat{d}_2$ of motion vector $(\hat{d}_1, \hat{d}_2)_{ij}$ respectively. The spatial continuity constrains of motion is implemented by median filter both on I_{hv} and I_{vv}. The result is reserved in matrixes I'_{hv} and I'_{vv} (see Equation (7)).

$$\begin{aligned}\boldsymbol{I'_{hv}[i,j]}&=\boldsymbol{mid}(\{\boldsymbol{I_{hv}[m,n]}|\boldsymbol{i}-\boldsymbol{W}\le\boldsymbol{m}\le\boldsymbol{i}+\boldsymbol{W},\boldsymbol{j}-\boldsymbol{W}\le\boldsymbol{n}\le\boldsymbol{j}+\boldsymbol{W}\})\\ \boldsymbol{I'_{vv}[i,j]}&=\boldsymbol{mid}(\{\boldsymbol{I_{vv}[m,n]}|\boldsymbol{i}-\boldsymbol{W}\le\boldsymbol{m}\le\boldsymbol{i}+\boldsymbol{W},\boldsymbol{j}-\boldsymbol{W}\le\boldsymbol{n}\le\boldsymbol{j}+\boldsymbol{W}\})\end{aligned} \tag{7}$$

where *W* is the radius of the median filter window and *mid(A)* returns the median element of set *A*.

Then, the temporal continuity constrains of motion is implemented further. Two successive motion vector images are considered in this paper. Let $I^{(k)}{}_v$ and $I^{(k+1)}{}_v$ be the motion vector images of frame *k* and *k+1* respectively. The temporal continuity constrains of motion is implemented then as Equation (8),

$$\begin{aligned}\boldsymbol{I''_{hv}[i,j]}&=\begin{cases}0 & \boldsymbol{if}(\left|\boldsymbol{I}^{(k)}{}_{\boldsymbol{hv}}\boldsymbol{[i,j]}-\boldsymbol{I}^{(k+1)}{}_{\boldsymbol{hv}}\boldsymbol{[i+I}^{(k)}{}_{\boldsymbol{vv}}\boldsymbol{[i,j],j+I}^{(k)}{}_{\boldsymbol{hv}}\boldsymbol{[i,j]]}\right|\ge\boldsymbol{MaxV})\\ \boldsymbol{I}^{(k)}{}_{\boldsymbol{hv}}\boldsymbol{[i,j]} & \boldsymbol{else}\end{cases}\\ \boldsymbol{I''_{vv}[i,j]}&=\begin{cases}0 & \boldsymbol{if}(\left|\boldsymbol{I}^{(k)}{}_{\boldsymbol{vv}}\boldsymbol{[i,j]}-\boldsymbol{I}^{(k+1)}{}_{\boldsymbol{vv}}\boldsymbol{[i+I}^{(k)}{}_{\boldsymbol{vv}}\boldsymbol{[i,j],j+I}^{(k)}{}_{\boldsymbol{hv}}\boldsymbol{[i,j]]}\right|\ge\boldsymbol{MaxV})\\ \boldsymbol{I}^{(k)}{}_{\boldsymbol{vv}}\boldsymbol{[i,j]} & \boldsymbol{else}\end{cases}\end{aligned} \tag{8}$$

where *MaxV* is a threshold, such as 6. The larger *MaxV* is, the stricter temporal continuity constrains of motion is supposed. Finally, the results of Equation (7) and (8) are integrated as Equation (9),

$$\boldsymbol{stmp[i,j]}=\begin{cases}1 & \boldsymbol{if}((|\boldsymbol{I'_{hv}[i,j]}|+|\boldsymbol{I'_{vv}[i,j]}|)\cdot(|\boldsymbol{I''_{hv}[i,j]}|+|\boldsymbol{I''_{vv}[i,j]}|)\neq 0)\\ 0 & \boldsymbol{else}\end{cases} \tag{9}$$

where *stmp* is a matrix. *Stmp*[*i,j*] equals to 1 if pixel (i, j) is judged as a moving pixel and 0 if background pixel.

4. MOVING OBJECT SEGMENTATION

Each frame has been segmented into a set of homogenous regions $\{R_1, R_2, ..., R_n\}$ in section II. It will be decided in this section that which of these regions belong to moving objects regions. This task is performed by integrating the result of image segmentation with moving pixels detection in section III. Those regions holding a certain ratio of moving pixels are decided to be moving objects regions. Let N_{R_i} be the number of pixels in region R_i. Region R_i is decided as a moving region if

$$\frac{\sum_{(i,j)\in R_i}\boldsymbol{stmp[i,j]}}{\boldsymbol{N}_{\boldsymbol{R_i}}}\ge\boldsymbol{T}_r \tag{10}$$

where T_r is a threshold, such as 40%. Then, a mask image can be gotten for each frame by Equation (11) to describe the moving objects regions in current frame.

$$\boldsymbol{mask[i,j]}=\begin{cases}1 & \boldsymbol{if(i,j)}\in\boldsymbol{R_i}\ \ \boldsymbol{and}\ \ \boldsymbol{R_i}\ \ \boldsymbol{is\ a\ moving\ region}\\ 0 & \boldsymbol{else}\end{cases} \tag{11}$$

The integration of image segmentation and moving pixels detection is a solution of the blank wall problem. If more than a certain ratio of pixels are judged as moving pixels all the pixels in the same homogeneous region are considered as moving pixels. Therefore, over-segmentation, which means many small homogeneous regions, goes against the solution of the blank wall problem. The comparison between Fig. 2(c) and Fig. 2(d) shows the effectiveness of the integration.

Post processing is used for better segmentation result. Morphological operators can be adopted to refine the border of moving objects regions. Then isolated small regions are deleted to get a clean mask. Finally, a clean object mask is gotten, which we can use to extract moving objects in a video sequence.

5. EXPERIMENTAL RESULTS

The proposed algorithm is applied in two test image sequences. The four parameters $\{\lambda$, *r, NUM1, NUM2*$\}$ in

the MSA are set as {10,0.1,100,10} in the experiments. The larger λ is, the lower the resolution of the segmentation is. The threshold r indicating the end of MSA. The number of iterations is relatively easy to be set, since the convergence of MSA has been verified.

There are three moving vehicles in the "Hamburg taxi" sequence shown in Fig. 2. It can be seen that the three moving vehicles are segmented accurately. To compare the proposed method with other existing approaches, Fig. 3 shows the results of two other region-based methods, image segmentation was performed by direct region growing, which leaded to deformed segmentation. Segmentation to a sequence of the real scene is illustrated in Fig. 4, it shows six segmented moving vehicles.

6. CONCLUSION

A new region-based algorithm combining gray image segmentation and moving pixel detection is presented in this paper for moving object segmentation on image sequence. Gray-scale image segmentation based on MSA can reduce over-segmentation in a noisy scene. Experimental results show that the method combining motion vector estimation by block-matching and spatio-temporal continuity constrains of motion is more efficient in detecting moving pixels than thresholding image difference.

7. ACKNOWLEDGEMENT

The work described in this paper was partially supported by the Hong Kong Polytechnic University under the Grant G-T228.

8. REFERENCES

[1] E. Cesmeil and D.L. Wang, "Motion segmentation based on motion/brightness integration and oscillatory correlation," *IEEE Transactions On Neural Network*, 11(4):935-947,2000.

[2] M. Bichsel, "Segmentation simply connected moving objects in a static scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(11):1138-1142,1994.

[3] C. Gu and M.-C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," *IEEE Transactions On Circuits And Systems For Video Technology*, 8(5):572-583,1998.

[4] J. Pan, S. Li and Y.-Q. Zhang, "Automatic extraction of moving objects using multiple features and multiple frames," *IEEE Int. Symposium on Circuits and Systems (ISCAS) 2000*, Geneva, 1:36 –39, 2000.

[5] D. Comaniciu and P. Meer, "Mean shift analysis, and applications," *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2 : 1197 –1203, 1999.

[6] L.J. Le Roux and J.J.D. Van Schalkwyk, "An overview of moving object segmentation in video images," *COMSIG 1991 Proceedings*, 53 –57, 1991.

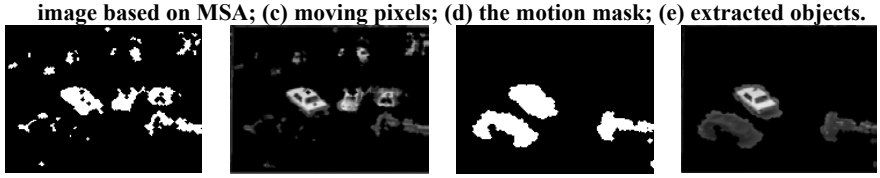
[7] K. Fukunaga and L.D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognitions," *IEEE Transaction On Information Theory*, 21:32-40,1975.

[8] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 17(8):790-799,1995.

[9] A. Neri, S. Colonnese, G. Russo and C. Tabacco, "Adaptive segmentation of moving object versus background for video coding," *SPIE*, 3164:443-453, 1997.

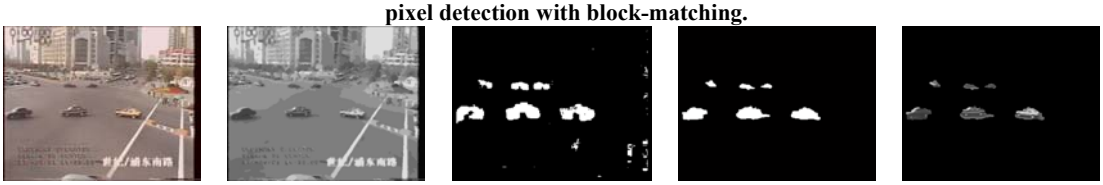
(a) (b) (c) (d) (e)

Figure 2: Segmentation on the ninth frame of "Hamburg taxi" image sequence. (a) original frame; (b) segmented image based on MSA; (c) moving pixels; (d) the motion mask; (e) extracted objects.



(a) (b) (c) (d)

Figure3: Segmentation with other methods. (a)-(b) segmentation based on MSA and moving pixel detection by thresholding image difference; (c)-(d) segmentation based on image segmentation with region growing and moving pixel detection with block-matching.



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

Figure 4: Segmentation of the image sequence of the Cross in Shanghai. (a) original frame; (b) segmentation based on MSA; (c) moving pixels; (d) the motion mask; (e) extracted objects.