Motion and Gray Based Automatic Road Segment Method MGARS in Urban Traffic Surveillance

Hong Liu, Jintao Li, Yueliang Qian, Shouxun Lin, and Qun Liu

Multilingual Interaction Technology and Evaluation Laboratory,
Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100080, China
hliu@ict.ac.cn

Abstract. This paper presents a novel method MGARS to automatic road area segmentation based on motion and gray feature for the purpose of urban traffic surveillance. The proposed method can locate road region by region growing algorithm with the fusion feature of motion information and grayscale of background image, which is independent to road marker information. An adaptive background subtraction approach using gray information is performed to motion segmentation. In region growing stage, start point that so called seed is selected automatically by motion centroid and local gray feature of background image. The threshold of region growing method is adaptively selected for different traffic scenes. The proposed method MGARS can effectively segment multi roads without manual initialization, and is robust to road surface pollution and tree shadow. The system can adapt to the new environment without human intervention. Experimental results on real urban traffic videos have substantiated the effectiveness of the proposed method.

1 Introduction

Automated visual traffic surveillance (AVTS) allows the visualization of vehicles on the road by using a single camera mounted in perspective view of the road scene that it is monitoring, thus enabling traffic-scene analysis [1]. In an AVTS system, moving vehicle detection is the basic task for other analysis. However, the performance of an AVTS system deteriorates when vehicles appear to occlude each other from the camera's point of view in a traffic video [2]. Failing to detect and resolve the presence of occlusion may lead to surveillance errors, including incorrect vehicle count, incorrect tracking of individual vehicles, and incorrect classification of vehicle type. As a result, methods for occlusion detection must be adopted in order to produce meaningful results [5]. These include stereo vision [6], an overhead camera with a viewing axis perpendicular to the road surface [7] or roadside mounted camera with a high position. Other researchers have done an extensive amount of work on occlusion detection and occlusion handling [5].

Occlusion problem is serious in urban traffic scenes for lower vehicle speed and little distance between vehicles than in highway scenes. In urban traffic monitoring, where camera is mounted roadside, occlusion is usually happened in far area. Instead of processing entire images, a computer vision system can analyze specific regions (the 'focus of attention') to identify and extract the features of interest [20]. So many papers

N. Zheng, X. Jiang, and X. Lan (Eds.): IWICPAS 2006, LNCS 4153, pp. 85 – 94, 2006.

propose to select better detect region with less vehicle occlusion. Tai [9] use detection line of each lane to detect whether the vehicle enters the detection region. Yu [10] use the Korean characters on each lane as the lane marks. Vehicle detection and shadow rejection are performed based on lane mark. But all the above detection region or detection lines are manually selected. In this case, the detection region is suited only to the current traffic video, which should be redefined for a new environment.

Region for vehicle detection can be seen as certain part of road area. In this paper, we focus on automatic road segment (ARS) approach, which is independent to any priori knowledge, such as road marker and camera viewing positions. Such a system would ease installation of the equipment due to its ability for self-initialization [11]. ARS is an important task for an adaptive traffic monitoring system. It enables the system to adapt to different environmental conditions.

In this paper, we propose a novel method MGARS for automatic road area segmentation in urban traffic video. The proposed method can locate road regions by the fusion feature of centroid of moving objects and gray of background image. The system block diagram is shown as Fig.1. An adaptive background subtraction approach using gray information is performed to motion segmentation. Then centroid of moving objects is obtained for next region growing process. Road regions are located using region growing method with automatically selecting seed points. The threshold of region growing method is adaptively selected for different traffic scenes. The proposed method can segment multi roads without manual initialization, and is robust to road surface pollution and tree shadow. Also it is independent to road marker information.

The rest of this paper is organized as follows. Section 2 provides a summary of previous work relevant to road segmentation in traffic video. The next three sections describe our proposed algorithm in details, and the results we obtained from experiments on a variety of traffic videos. Section 6 concludes by describing some of the important characteristics of our algorithm, and directions for future research.

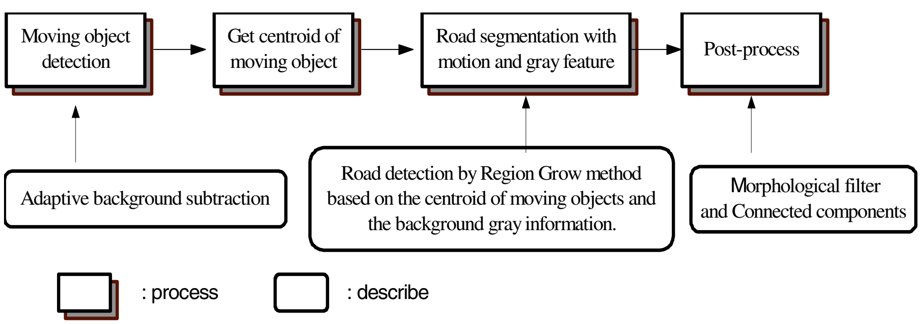


Fig. 1. Block diagram of automatic road segment method MRARS

2 Previous Work

In this section, we focus on road segment method in previous research. Real-time road segmentation is complicated by the great variability of vehicle and environmental conditions. Changing seasons or weather conditions, time of the day, dirt on the road, shadows. Because of these combined effects, robust segmentation is very demanding. The approaches in lane detection can be distinguished into two classes, namely lane-region detection and lane-border detection [4].

Lane-region approaches detect the lane with the changing intensity distribution along the region of a lane. For automatic vehicle guidance case, [18] assumes the road just in front of car, and take a few sample to capture the road color. Then flood-fill the road region using the sampled colors. The lane-region analysis can also be modeled as a classification problem, which labels image pixels into road and non-road classes based on particular features, which required extensive training process. Ref. [12] uses Gaussian distributions of (R,G,B) values to model the color classes. Ref. [13] use the hue, saturation, gray-value (HSV) space as more effective for classification. Besides color, the local texture of the image has been used as a feature for classification. Ref. [16] uses a normalized gradient measure based on a high-resolution and a low-resolution (smoothed) image, in order to handle shadow interior and boundaries. However changes in outdoor illuminations may change the road colors perceived by the camera and introduce errors in the classification. Ref. [11] uses motion information in lane detection. An activity map is used to distinguish between active areas of the scene where motion is occurring (the road) and inactive areas of no significant motion. But lane finding will false when the vehicles change their lane.

Lane-border detection method considers directly the spatial detection of lane characteristics. According the difference of lane characteristics, two general subclasses involve feature-driven approaches and model-driven approaches [4]. Feature-driven approaches are based on the detection of edges in the image and the organization of edges into meaningful structures (lanes or lane markings). The Road Markings Analysis (ROMA) system is based on aggregation of the gradient direction at edge pixels in real-time [14]. In general, edge feature suffer from noise effects, such as strong shadow edges sometime. The aim of model-driven approaches is to match the road edges with a deformable template, which is usually used in vehicle guidance. The Hough Transform is used to extract road boundaries from an image [17]. Ref. [15] use snakes to model road segments. Model-based approaches for lane finding have been extensively employed in stereo vision systems. Such pproaches assume a parametric model of the lane geometry, and a tracking algorithm estimates the parameters of this model from feature measurements in the left and right images. Model-driven approaches provide powerful means for the analysis of road edges and markings. However, the use of a model has certain drawbacks, such as the difficulty in choosing and maintaining an appropriate model for the road structure, the inefficiency in matching complex road structures and the high computational complexity [4].

3 Robust Motion Segmentation

3.1 Gray Based Background Subtraction

In recent years time-adaptive per pixel mixtures of Gaussians background models have been a popular choice for modeling complex and time varying backgrounds [6].

This method has the advantage that multi-modal backgrounds (such as moving trees) can be modeled. Different to Stauffer's method [19], we use only gray value of source image to construct background image.

In [19], each pixel is modeled as a pixel process; each process consists of a mixture of k adaptive Gaussian distributions. The distributions with least variance and l maximum weight are isolated as the background. The probability that a pixel of a particular distribution will occur at time t is determined by:

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \ . \tag{1}$$

where K is the number of Gaussian distributions, $\omega_{i,t}$ is the weight estimate of the ith Gaussian in the mixture at time t, $\mu_{i,t}$ and $\sum_{i,t}$ are the mean value and covariance matrix of the ith Gaussian at time t, and η is the Gaussian probability density function.

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{\pi}{2}} \mid \Sigma \mid^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)}, \ \Sigma_{k,t} = \sigma_k^2 I \ . \tag{2}$$

In this paper we set k=3 Gaussians. An on-line k-means approximation algorithm is used for the mixture model. Every new pixel X_t is checked against the K existing Gaussian distribution. A match is found if the pixel value is within L = 2.5 standard deviation of a distribution. This is effectively per pixel per distribution threshold and can be used to model regions with periodically changing lighting conditions.

If the current pixel value matches none of the distributions the least probable distribution is updated with the current pixel values, a high variance and low prior weight. The prior weights of the K distributions are updated at time t according to:

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \ . \tag{3}$$

where α is the learning rate and $M_{k,t}$ is 1 for the model which matched the pixel and 0 for the remaining models. We set learning rate α =0.002. The changing rate in the model is defined by $1/\alpha$. That is means after 500 frames the background model well updated fully. After this approximation the weights are renormalized, the parameters μ and σ for the unmatched distributions remain the same. The parameters for the matching distribution are updated as follows:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t \ . \tag{4}$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) . \tag{5}$$

where
$$\rho = \alpha\eta(X_t \mid \mu_k, \sigma_k) . \tag{6}$$

For change detection a heuristic searches for the learnt distributions that have more supporting evidence. The Gaussians are ordered based on the ratio of ω/σ. This increases as the Gaussian's weight increases and its variance decreases. The first B distributions accounting for a proportion T of the observed data are defined as background. We set T=0.8 in this paper.

$$B = \arg\min_b (\sum_{k=1}^{b} \omega_k > T) \ . \tag{7}$$

For the non-background pixel, we calculate the gray difference between this pixel in current image and in background model. Only the pixel with the difference over the threshold 10 is labeled as foreground pixel or motion pixel. Then the difference image is binary. In our experiments, the background model can be usually obtained within 100 to 200 sequence frames, which is good enough for the future segmentation process.

3.2 Obtain Centroid of Moving Object

After motion segmentation, the moving objects with their bounding boxes and centroids are extracted from each frame. The centroid (x,y) of a bounding box B corresponding to an object O is defined as follows:

$$x = (\sum_{i=0}^{N} x_i) \Big/ N \, , \, y = (\sum_{i=0}^{N} y_i) \Big/ N \ . \tag{9}$$

where N is the number of pixels belong to object O within bounding box B, x_i and y_i represent the x-coordinate and y-coordinate of the *i*th pixel in object O.

Fig.2 shows some results on an urban traffic video. Fib.2a is one source frame. The constructed background image is showed in Fig.2b, which use 100 sequence frames with moving objects. And the motion segment result on source image and binary image are presented as Fig.2c and Fig.2d. The motion centroids with 3×3 region after processing 200 sequence frames are show in Fig.2e.

Fig. 2a. Source image

Fig. 2b. Constructed background image

Fig. 2c. Motion segment

Fig. 2d. Binary image

Fig. 2e. Motion centroids

4 Motion and Gray Based Automatic Road Segmentation MGARS

In this section, we describe the proposed automatic road segment method using the feature of motion centroids and gray value of background image.

First a 3×3 Gaussian filter is performed on background image to reduce noise. Experimental results will show the effective of this smooth process.

The canny edge detection algorithm is known as the optimal edge detector. We use canny edge detection on background image to get edge information.

The region grow algorithm that is also called flood fill method is used in our system. Region growing process starts with some point, called "seed", fills the seed pixel neighborhoods inside which all pixel values are close to each other. This process is propagated until it reaches the image boundary or cannot find any new pixels to fill due to a large difference in pixel values.

This algorithm need some parameters: coordinates of the seed point inside the road area, threshold τ as maximal lower difference and maximal upper difference between the values of pixel belonging to the filled domain and one of the neighboring pixels to identify, type of connectivity. If connectivity is four, the region growing process tries out four neighbors of the current pixel otherwise the process tries out all the eight neighbors.

Ref. [18] uses flood fill method to extract the road region in front of moving vehicle in vehicle guidance. The seed is selected by priori knowledge, which make sure the seed must be in the road region. And the threshold is a constant value, which may not adapt to different scene.

In this paper, we propose a novel method MGARS to segment road regions automatically. The method consists of following stages:

1. Gray background image is divided into 10×10 pixel partitions without overlap.
2. For each partition, the number of moving centroid, mean and standard deviation of gray value and number of edge pixel are calculated. Partition containing motion centroids is called Centroid Partition here. The total number of Centroid Partition can also be calculated together.
3. From the bottom of background image, we search the proper Centroid Partition, which have number of motion centroids more than 2, standard deviation less then 10 and the number of edge pixel less than 20. Then the center of this Centroid Partition is selected as seed point for region growing process.
4. Process 8-connectivity region growing algorithm from the above seed point. If the gray value of neighbor pixel is similar to the gray value of seed point according to the threshold τ, this neighbor pixel is filled.
5. If 30 percent of pixels in a partition is filled, we define this Centroid Partition is filled. Calculate how many Centroid Partitions are filled. If there is enough Centroid Partitions (90 percent in our system) are filled by region growing process, the seed point search process can stop, else back to stage 3 to search the next proper seed point for region growing.
6. If the whole partitions in background image are searched, the region growing process can stop.

In stage 3, the strategy we choose proper Centroid Partition is considering the seed point should in the road region, which can be got by centroid of moving objects, and its neighbor region should relatively smooth. The method can select seed point avoid noise pixel, such as shadow, edge and road smut.

In stage 4, the min and max difference that is threshold τ is selected automatically. Here, we calculate the mean and standard deviation of all Centroid Partitions. If the standard deviation is less than 20, which means gray value of road region is so smooth, the threshold τ is set to 1, else the threshold τ is set to 2. The compare experiment is show in next section.

After region growing process, most of the road pixel is filled, then binary the result image. Post-processing stage using morphological filter is performed on binary image to remove small region and smooth the boundary of road region. Then road regions are labeled as connected components.

5 Experiments

The test video sequences were taken using a camera on roadside or cloverleaf junction in urban. The video was sampled at a resolution of 320×240 and a rate of 25 frames per second. We used only grays value of source video, and output a grayscale background model. Tests were performed on two representative sequences that might be commonly encountered in urban surveillance.

τ =1

Fig. 3a. Background image with seed position

Fig. 3b. Result of region grow

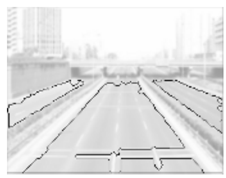


Fig. 3c. Edge of extracted road region

τ =2

Fig. 3d. Background image with seed position

Fig. 3e. Result of region grow

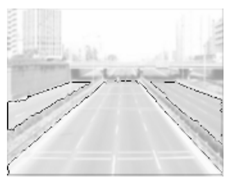


Fig. 3f. Edge of extracted road region

τ =3

Fig. 3g. Background image with seed position

Fig. 3h. Result of region grow

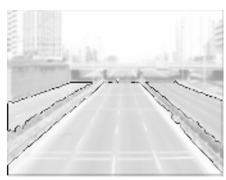
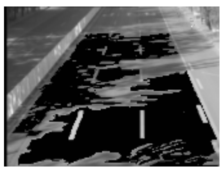


Fig. 3i. Edge of extracted road region

The first traffic scene with multi roads is medium shot. Fig.3 illustrate the road segment results using our proposed method. The first row in Fig.3 shows the results with threshold τ =1 in region growing process. Fig.3a displays the positions of proper Centroid Partition with seed point. Fig.3b is the result after region growing process from the selected seed point as in Fig.3a. After morphology filter, superimposition of the edges of road region onto the lighter original background image is shown in Fig.3c. It also demonstrates improvements possible using morphological filters in a post-processing stage. The results with threshold τ =2 and τ =3 are shown as the second and the third row of Fig.3. From the experiment results, we can conclude that if the threshold τ is lower, there will need more seed point to flood fill the whole road regions. But as Fig.3c shows lower threshold can not get well segment result if road region is non consistency in grayscale. The threshold τ =2 can get similar better result compared with threshold τ =3. The marker on the road region can be removed by post-process as shown in Fig.3f. In our system, for this video, threshold τ is automatically selected as 2 according to the standard deviation of all the Centroid Partitions. This experiment can also illustrate that our algorithm can detect multi road regions effectively.

τ =1



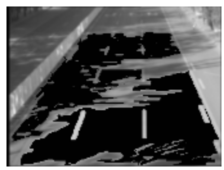
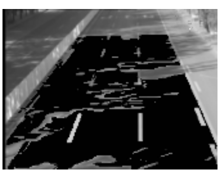


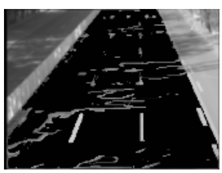
Fig. 4a. Background image with shadow

Fig. 4b. Result of region grow without smooth

Fig. 4c. Result of region grow with Gaussian smooth

τ =2





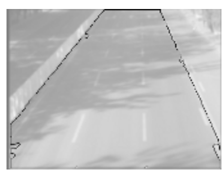


Fig. 4d. Result of region grow

Fig. 4e. Result of region grow with Gaussian smooth

Fig. 4f. Edge of extracted road region

A typical traffic surveillance scene in urban was considered next where the challenge was due to the vigorous motion of the trees and the strong shadow on road surface. And the right part is pavement with moving person or bicycle. The result is shown in Fig.4. Fig.4a is the background image constructed, which has tree shadow on the road surface. Fig.4b is the result of flood filling by threshold τ =1 and without Gaussian filter on background image. While Fig.4c is the result by threshold τ =1 and after Gaussian filter, which can fill more road pixels than Fig.4b. Fig.4d is the result by threshold τ =2 without Gausssian filter while Fig.4e is by threshold τ =2 and after Gaussian filter. From the segment result, we can see that Gaussian filter on the source background image before region growing can get better segment results for it can smooth background image and remove many noise. Fig.4f is the final edge of road region after morphological filter. By the way, the left road is not extracted here for there is no moving vehicles appeared in 200 frames in which we get motion centroids. If we use more frames to get moving objects, moving vehicle will appear on the left road and it can also be extracted well.

In the experiment, our algorithm successfully segments out the entire road regions containing moving vehicles.

6 Discussion and Conclusions

In this paper, we present a novel method MGARS to automatically extract road region from traffic monitoring video. Fusion features with moving segmentation and gray of background image are used for region growing process to get road region. An adaptive background subtraction method is proposed and applied to several real life traffic video sequences to obtain more accurate motion information of the moving objects. Road regions are located using region growing method with automatically selecting seed by motion centroid and local gray feature of background image. The threshold of region growing method is adaptively selected for different traffic scenes. Satisfactory results were obtained in our experiment, and all the road areas with moving vehicles are successfully identified through our algorithm. The method shows robustness to variations in both road properties and illumination conditions. The algorithm is viewpoint independent. No manual initialization or prior knowledge of the road shape is needed, the method can also suit to curve road. The proposed method can detect multi roads together, and can adapt to the new environment without any human intervention.

Several problems occurred in the experiments. In the stage of motion segmentation, moving shadow is detected as moving object and the presence of occlusions between vehicles make centroid position of moving object is not the real centroid of moving vehicle. Also in some case, moving object contains moving person or bicycle, which will disturb the result of moving vehicle segmentation. Future improvements include using techniques that involve modeling of the motion of vehicles and pedestrians in order to produce a better classier. Moving shadow detect can also considered in the future research. Further, since the position of the centroid of a moving vehicle is recorded during the segment process, this information can be used in the future for extracting moving trajectory.

Acknowledgment

The research is sponsored by National Hi-Tech Program of China (No. 2004AA-114010, 2003AA111010).

References

1. Beymer, D., Malik, K.: Tracking Vehicles in Congested Traffic. Proc. IEEE Intelligent Vehicles Symp. (1996) 130–135
2. Yung, N. H. C., Lai, A.H.S.: Detection of Vehicle Occlusion Using a Generalized Deformable Model. Proc. IEEE Int. Symp. Circuits and Systems, 4 (1998) 154–157
3. Fathy, M., Siyal, M. Y.: A Window-based Image Processing Technique for Quantitative and Qualitative Analysis of Road Traffic Parameters. IEEE Trans. On Vehicle Technology, (1998) 1342–1349
4. Kastrinaki, V., Zervakis, M., Kalaitzakis, K.: A Survey of Video Processing Techniques for Traffic Applications. Image and Vision Computing 21(2003) 359-381
5. Clement, C.C.P., William, W.L.L., Nelson, H.C.Y.: A Novel Method for Resolving Vehicle Occlusion in a Monocular Traffic-Image Sequence. IEEE Trans. On Intelligent Transportation Systems, 5 (2004) 129-141
6. Rabie, T., Shalaby, A., Abdulhai, B., El-Rabbany, A.: Mobile Vision-based Vehicle Tracking and Traffic Control. Proc. IEEE 5th Int. Conf. Intelligent Transportation Systems, (2002) 13–18
7. Lai, A. H. S.: An Effective Methodology for Visual Traffic Surveillance. Ph.D. dissertation, Univ. Hong Kong, China, 2000
8. Ikeda, T., Ohnaka, S., Mizoguchi, M.:Traffic Measurement with a Roadside Vision System-Individual Tracking of Overlapped Vehicles. Proc. IEEE 13th Int. Conf. Pattern Recognition, 3 (1996) 859–864
9. Jen-Chao, T., Shung-Tsang, T.: Real-time Image Tracking for Automatic Traffic Monitoring and Enforcement Applications. Visual tracking. Image and Visison Computing, 22(2004) 640-649
10. Yu, M., Jiang G.Y., He S.L.: Land Mark Based Method for Vehicle Detection and Counting from Video. Chinese Journal of Scientific Instrument, 23(2002) 386-390
11. Stewart, B.D., Reading, I., Thomson, M.S., Binnie, T.D., Dickinson, K.W., Wan, C.L.: Adaptive Lane Finding in Road Traffic Image Analysis, Proc. of Seventh International Conference on Road Traffic Monitoring and Control, IEE, London (1994)
12. Thorpe, C., Hebert, M.H., Kanade, T., Shafer, S.A.: Vision and Navigation for the Carnegie–Mellon Navlab. IEEE Trans. on Pattern Analysis and Machine Intelligence 10 (3) (1988)
13. Betke, M., Haritaoglu, E., Davis, L.S.: Real-time Multiple Vehicle Detection and Tracking From a Moving Vehicle. Machine Vision and Applications 12 (2000) 69–83
14. Enkelmann, W., Struck, G., Geisler, J.: ROMA—a System for Model-based Analysis of Road Markings. Proc. of IEEE Intelligent Vehicles, Detroit (1995) 356–360
15. Yuille, A.L., Coughlan, J.M.: Fundamental limits of Bayesian Inference: Order Parameters and Phase Transitions for Road Tracking. IEEE Pattern Analysis and Machine Intelligence 22 (2) (2000) 160–173
16. Taylor, C.J., Malik, J. Weber, J.: A Real Time Approach to Stereopsis and Lane-Finding, IFAC Transportation Systems Chania, Greece (1997)
17. He, Y. H., Wang, H., Zhang B.: Color Based Road Detection in Urban Traffic Scenes, IEEE Trans. on Intelligent Transportation Systems. 5(2004) 309-318
18. Park, E., Tran, B., Arfvidsson, J.: Freeway Car Detection. CS 223B, Stanford, January 25, 2006
19. Stauffer, C., Grimson, W.: Adaptive Background Mixture Models for Real-Time Tracking. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, (Fort Collins, CO), (1999)246–252
20. Bertozzi, M, Broggi, A., Fascioli, A.: Vision-based Intelligent vehicles: State of Art and Perspectives. Robotics and Autonomous System 32(2000) 1-16