Urban road area recognition in ITS based on mean shift method

Zhaoxue Chen (陈兆学) and Pengfei Shi (施鹏飞)

Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030

Received May 30, 2003

A color-based visual technique is described based on the mean shift image segmentation method providing relevant information for robust localization of the visible road area in Urban Intelligent Transportation System (U-ITS). The traffic image sequences are firstly trained to extract the background and then segmented into separated parts by the mean shift method as initialization, regions with the number of pixels not less than a threshold and with more uniform surfaces with the "same" color compared to their environment are filtered as recognized road area. The algorithm given in this paper can present road area recognition with arbitrary shapes, which is fit for unstructured road applications in urban cities very well.

OCIS codes: 100.0100, 100.2960.

Digital video data is often gathered with certain resolution and qualities by video cameras installed on a link or an intersection for static or dynamic surveillance and management in Urban Intelligent Transportation System (U-ITS). Different traffic information can be extracted for various purposes such as enforcement (e.g. license plate recognition), surveillance (e.g. traffic light status and vehicle detection, degree of saturation computation for lanes and simple video monitoring), incident management (incident detection and avoidance) and public video information distribution service.

The road area segmentation of traffic images is a crucial step for analysis of the complex urban traffic scenes. Many traditional systems have been proposed to discover parts of the traffic scenes by extracting geometry features to infer and verify the existence of certain classes of objects. Another relevant class of segmentation methods is based on region growing and clustering. In this project, a color-based visual technique is described based on the mean shift image segmentation method providing relevant information for robust localization of the visible road area independently even with the presence of different lighting conditions. The traffic image sequences are firstly trained to extract the background and then segmented into separated parts by the mean shift method as initialization, regions with the number of pixels not less than a threshold and with more uniform surfaces with the "same" color compared to their environment are filtered as recognized road area. The algorithm can present road area recognition with arbitrary shapes, which is fit for unstructured road applications in urban cities very well.

The road area is relatively static with few changes in traffic scenes during surveillance, but is easily occluded by dynamic foreground objects such as vehicles and people, which influence the road area segmentation efficiency greatly. Because the vehicle volume is large in urban traffic scenes, it is difficult to filter an available view with enough unobstructed road area in many cases. To avoid disturbance from presence of foreground objects, before the road area recognition, a background training and extraction from video sequences are necessary as a pre-processing step at first.

At present, there has been a large amount of work addressing the issues of background in video surveillance field. The algorithm exploited in this paper has been already presented in Ref. [1], which is efficient for stronger background initialization for adoption of low-level motion optic information and some heuristics. Briefly-speaking, the algorithm takes a video sequence as input and produces a statistical background model describing the static parts of the scene, which are just what we exactly need in this paper. At first multiple hypotheses of the background value at each pixel are generated by locating periods of stable intensity in the sequence, and then the likelihood of each hypothesis is evaluated using optical flow information from the neighborhood around the pixel, and the most likely hypothesis is chosen to represent the background. Because background

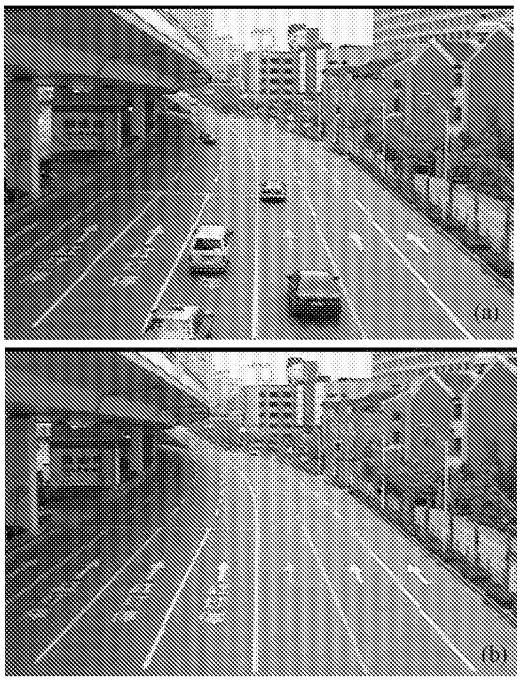

Fig. 1. (a) A snapshot of traffic video sequence; (b) obtained background picture (352 × 223 image).

extraction is not the concern in this paper, for detailed description of this algorithm, please refer to Ref. [1] directly. The author has extended the algorithm in this paper to use color image sequences and output a color background model for the subsequent work.

A snapshot of training sequence and the obtained colorful traffic background are given in Fig. 1.

A large class of image segmentation algorithms are based on feature space analysis. Given an image, feature vectors are extracted from local neighborhoods and mapped into the space spanned by their components. Significant features in the image then correspond to high-density regions in this space. Feature space analysis is therefore the procedure of recovering the centers of the high-density regions, i.e., the representations of the significant image features. In the traditional clustering techniques the feature space is modeled as a mixture of multivariate normal distributions, which can introduce severe artifacts due to the elliptical shape imposed over the clusters or due to an error in determining their number. As a nonparametric density estimation with no use of the normal assumption, the mean shift can eliminate these artifacts efficiently and present robust image segmentation in arbitrary shapes by the associated iterative procedure of mode seeking. The mean shift procedure is an extremely versatile tool for feature space analysis and provide reliable solutions for many vision tasks.

Let $\{x_i\}_{i=1\cdots n}$ be an arbitrary set of n points in the d-dimensional Euclidean space R^d, x be the center of a d-dimensional kernel, which is given by $K(x)$. The kernel density estimate for $K(x)$ is given by

$$\hat{f}_K(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \tag{1}$$

where $\hat{f}_K(x)$ is the kernel density function for $K(x)$, and h is the radius of the kernel. Possible kernels include uniform kernel, Gaussian kernel, triangular kernel, biweight kernel and Epanechnikov kernel. The quality of a kernel density estimator is measured by the mean of the square error between the density and estimate, integrated over the domain of definition. In practice, however, only an asymptotic approximation of this measure (denoted as asymptotic approximation minimum mean integrated square error (AMISE)) can be computed. Under the asymptotics, the number of data points $n \to \infty$, while the bandwidth $h \to 0$ at a rate slower than n^{-1}. Among these kernels, the Epanechnikov kernel yields minimum AMISE measure[3]. Therefore we used this kernel to calculate the kernel density estimate of the data that lies within the kernel. d-dimensional Epanechnikov kernel is given by

$$K_E(x) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1 - x^T x) & x^T x < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where c_d is the volume of the unit d-dimensional sphere[2]. Then, the mean shift vector computed at location x is given by[4]

$$M_h(x) = \frac{h^2}{d+2}\frac{\nabla \hat{f}_K(x)}{\hat{f}_K(x)}, \tag{3}$$

where $\nabla(\cdot)$ denotes gradient operator. It can be shown in Eq. (3) that the mean shift vector, i.e. the difference between local average value and the center of searching window, has the direction of the gradient of the density estimate at x when this estimate is obtained with the Epanechnikov kernel. Since the mean shift vector always points towards the direction of the maximum increase in the density, the successive computation of Eq. (3), followed by the translation of the kernel $K_E(x)$ by $M_h(x)$ will define a path that converges to a local maximum of the underlying density, i.e., to a mode of the density. This algorithm is called the mean shift iterative procedure for mode in feature space.

For efficient image segmentation, the mean shift procedure can be applied for the data points in the joint spatial-range domain[4], where the space of the 2-dimensional lattice represents the spatial domain, with kernel radius h_s, and the space of 3 color components constitutes the range domain, with kernel radius h_r. In this approach, a data point defined in the spatial-range domain is assigned with a point of convergence, which represents the local mode of the density in this space. One can define displacement vector in the spatial domain as the spatial difference between convergence point and the original point. The convergence points sufficiently closed in the joint 5-dimensional feature space for color level images are gathered together to form uniform regions of image segmentation.

In the urban traffic image, road area has usually such characteristics as follows: (1) Regions with enough number of pixels in the image are considered as road areas. (2) Urban road usually has an expected structure with several lanes. (3) Road areas have quasi-uniform color. (4) Road continuity may be interrupted for different reasons such as objects on road and shadows projected from trees or building. Based on these features of the urban road area, algorithms to recognize road area are discussed below.

At first, the traffic image is segmented into separated parts by the mean shift method with a smaller kernel window size $(h_s, h_r) = (7, 3)$ to show more details as initialization, as shown in Fig. 2[3], then regions with the number of pixels not less than a threshold are filtered to get original candidates of road area. At last, considering the road areas generally have more uniform surfaces with the "same" color compared to its environment, the road area can be effectively recognized.

In urban traffic image, the environments are complex

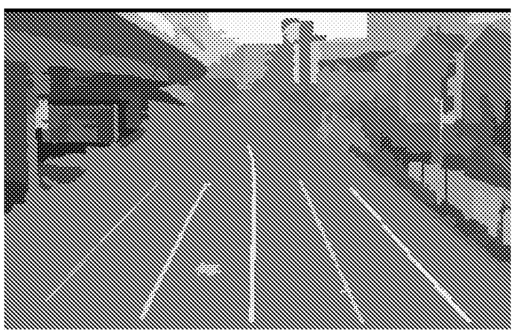

Fig. 2. Initialization of road area recognition by mean shift with $(h_s, h_r) = (7, 3)$.

and road areas may be extended to the whole view of the video camera and it is very hard to get initial color sample of the road area. Because the road areas usually have lanes within them, in order to get much better road color samples, the author analyzes the shape of the regions at first from mean shift process to get part of lanes and sample the road color around lane areas. During this step, the author found the so-called region compactness was interesting and useful. Definitely, compactness c can be defined as

$$c = \frac{\mu}{\sigma}, \mu = \frac{1}{N}\sum_{n=0}^{N-1} \|(x_n; y_n) - (x_c; y_c)\|, \quad (4)$$

$$\sigma^2 = \frac{1}{N}\sum_{n=0}^{N-1} (\|(x_n; y_n) - (x_c; y_c)\| - \mu)^2, \quad (5)$$

where $(x_c, y_c) = r_c$ is the vector of the centroid coordinates of a region. Following the increase of the value of c, the region has a shape similar to a circle (the most compact shape of all). Therefore irregular shapes, for example those greatly elongated in a particular direction give rise to very small values of c. Because lanes are always elongated with white color, it is very easy to recognize part of them based on c value associated with color and intensity information.

The HSV (hue, saturation, and value) color space was employed since its metric is a satisfactory approximation to Euclidean, thus allowing a better control over variations in pixel values for the same color. Since image region around lanes can be considered as road areas, a small square (20×20 pixels) is used as a road color template (test area). Then, compute the mean value and standard deviation over hue, saturation, and intensity value of the test area. The estimated HSV value of road is within the range (mean −5*std, mean +5*std) for the experiment. The mean HSV value of each road area candidates of the image is compared with this estimated value, and regions within this range are considered as road areas. Figure 3 shows the result of the whole recognized road area.

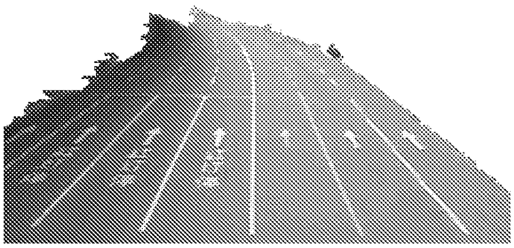

Fig. 3. Result of the whole recognized road area.

In conclusion, a new algorithm for road area recognition in U-ITS based on mean-shift method was presented in this paper, which can be used in traffic surveillance as an essential step. An urban traffic scene is very complex for presence of many artificial buildings and different foreground objects in the monitor view, which cause great disturbance to various traffic parameter extraction algorithms such as vehicles detection on road and degree of saturation (DOS) computation etc. Because of the road area is almost keeping static with few changes during the whole surveillance process, committing road area recognition beforehand could simplify the following algorithms for extracting various traffic information greatly.

Based on mean-shift method, the algorithm given in this paper presents robust road area recognition with arbitrary shapes, which are fit for unstructured road scene in urban cities very well. The nearer future work is to develop accurate lanes detection method and model approximation based on the presented algorithm as preparation for various video traffic analysis. For further improvement, it can even be applied in different fields such as smart car or intelligent robot navigation.

This work was supported by the National Key Project for Basic Research on Urban Traffic Monitoring and Management System (PRA SI01-01, G1998030408). Z. Chen's e-mail address is chenzhaoxue@sjtu.edu.cn.

References

1. D. Gutchess, M. Trajkovics, E. Cohen-Solal, D. Lyons, and A. K. Jain, in *Proceedings of Eighth IEEE International Conference* **1,** 733 (2001).
2. D. Comaniciu, V. Ramesh, and P. Meer, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2,** 142 (2000).
3. D. Comaniciu and P. Meer, IEEE Trans. Pattern Anal. Machine Intell. **24,** 603 (2002).
4. D. Comaniciu and P. Meer, in *IEEE International Conference on Computer Vision* 1197 (1999).
5. K. Liu, Y. X. Liu, J. Y. Yang, Y. Q. Cheng, and N. C. Gu, Proc. SPIE **1613,** 340 (1992).
6. Y. Z. Cheng, IEEE Trans. Pattern Anal. Machine Intell **17,** 790 (1995).
7. I. Y. H. Gu and V. Gui, in *Proceedings of IEEE International Conference on Image Processing* **1,** 726 (2001).
8. D. Comaniciu and P. Meer, in *Proceedings of IEEE Computer Society Conference* 750 (1997).
9. H. Cheng, N. N. Zheng, and T. Liu, in *Signal Processing of 2002 6th International Conference* **2,** 1059 (2002).