

Momentum Based Level Set Method For Accurate Object Tracking

Haocheng Le¹, Linglong Hu², Yuanjing Feng³

Institute of Information Processing and Automation, College of Information Engineering,
Zhejiang University of Technology
Zhejiang Provincial United Key Laboratory of Embedded System
Hangzhou, Zhejiang, 310023, China

1)lhc7895123@sina.com 2)linglonghu@126.com 3)fyjing@zjut.edu.cn

Abstract—This paper proposes a novel object tracking method that is robust to a cluttered background and large motion. First, a posterior probability measure (PPM) is adopted to locate the object region. Then the momentum based level set is used to evolve the object contour in order to improve the tracking precision. To achieve rough object localization, the initial target position is predicted and evaluated by the Kalman filter and the PPM, respectively. In the contour evolution stage, the active contour is evolved on the basis of an object feature image. This method can acquire more accurate target template as well as target center. The comparison between our method and the kernel-based method demonstrates that our method can effectively cope with the deformation of object contour and the influence of the complex background when similar colors exist nearby. Experimental results show that our method has higher tracking precision.

Index Terms—posterior probability measure, Kalman filter, momentum, level set, object tracking

I. INTRODUCTION

Object tracking is a crucial task within the computer vision field. It has been widely applied in applications such as video surveillance [1], perceptual user interface [2], video coding [3] and driver assistance [4]. Various tracking algorithms have been proposed in these years [5-10], for ease of discussion, which are divided into two categories [6,10], namely, region-based method and contour-based method.

The basic idea of region-based method is to track object with the similarity measure of object region. The most widely used similarity measures include the Bhattacharyya coefficient [11-13], Kullback-Leibler divergence [12,14] and so on, and the mean-shift algorithm in which uses Bhattacharyya coefficient has achieved considerable success in image segmentation and tracking [15-19], for its simplicity and robustness. The real-time kernel-based object tracking proposed by Comaniciu [15] can successfully track partial occluded nonrigid objects, but cannot cope with the deformation of object contours and will cause the serious bias when similar colors exist nearby. A more discriminative similarity measure in spatial-feature space was proposed by Yang [20], which can effectively cope with the

translation and scaling of the object, but does not consider the rotation invariance. During a visual tracking process, the scale of the tracked target may continuously vary. In order to achieve precise localization, the target scale or the kernel bandwidth must be adaptive. Numerous methods have been developed for adaptive adjustments of the target scale. Comaniciu et al. [16] proposed to apply the mean-shift procedure three times with three different kernel bandwidths derived by adding or reducing 10 percent of the current target size during each repetition, from which the optimal scale with the maximum Bhattacharyya coefficient will be selected. A data-driven scale selection method was presented in Refs. [21,22]. By analyzing the local statistical characteristics around each data point, the kernel bandwidth is calculated. Collins developed a scheme for blob tracking over the scale space [23]. A scale space mean-shift procedure is first executed and then the spatial space calculation is processed over multiple scales. But all of the scale updating methods cannot solve the object deformation very effectively.

For contour-based methods, Algorithm can be performed using two different approaches. The first approach uses state space models to model the contour shape and motion. The second approach directly evolves the contour by minimizing the contour energy using direct minimization techniques such as gradient descent. Terzopoulos and Szeliski in 1992 define the object state by the dynamics of the control points, which modeled in terms of a spring model moving the control points based on the spring stiffness parameters. Chen et al. in 2001 propose a contour tracker where the contour is parameterized as an ellipse, in which each contour node has an associated HMM and the states of each HMM is defined by the points lying on the lines normal to the contour control point. While another approach, Yilmaz [24] incorporated prior shape into object energy functions, and used level set to evolve the contour by minimizing the energy functional. Chung and Chen [25] presented a video segmentation system that integrated Markov random field (MRF)-based contour tracking with graph-cut image segmentation. Contour-based methods can achieve a high tracking precision, but their robustness is usually not better than that of region-based methods. Furthermore, the computing cost of contour-based

methods is usually high, especially for large and fast-moving objects.

Combining the merits of region-based and contour-based methods, we introduce a robust object tracking method. First, the PPM is adopted to locate the object region, and the Kalman filter is used to determine the initial object tracking position. Then the object feature image is generated after a series of pre-processing, and the momentum based level set segmentation is used to evolve the object contour in order to improve the tracking precision.

The paper will proceed as follows. In Section II, we describe the idea of target prediction with the Kalman filter and rough object localization with PPM. Then, Section III presents how to evolve the object contour in the object feature space with the momentum based level set segmentation. This method will be tested and verified in Section IV, and simultaneously we will give out our results of experiments. Finally, Section V concludes the paper and presents ideas for our future work.

II. TARGET LOCALIZATION

A. Target Prediction Based on the Kalman Filter

The Kalman filter was proposed by R.E. Kalman with his famous paper in 1960, describing a recursive solution to the discrete-data linear filtering problem. It consists of a set of mathematical equations that provides an efficient computational means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very useful in three aspects: estimations of past, present, and even future states. In this paper, we use the Kalman filter to predict the center $[x_0, y_0]$ of the object region before beginning to track. Let v_x and v_y represent the velocities in the x and y directions, respectively, and the state vector can be represented as $x = [x_0, y_0, v_x, v_y]$. Therefore the system can be modeled as

$$x_{t+1} = \phi x_t + w_t. \quad (1)$$

The measurement model in the form needed by the Kalman filter is

$$z_t = Hx_t + v_t. \quad (2)$$

where

$$\phi = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Given the state model in (1) and measurement model in (2) as well as some initial conditions, the state vector x_{t+1} can be updated using the system model (for prediction) and measurement model (for updating). More details about the Kalman filter can refer to [26] and [27].

B. Object Tracking with the PPM

Image matching is an important part for visual tracking, and the Bhattacharyya coefficient usually is an efficient method. But since the influence of background feature, the optimal location obtained by Bhattacharyya coefficient may not be the exact target location. Thus, biased or even wrong location may occur in visual tracking. Taking into account the statistical features of a search region, PPM can decrease the interference of the background pixels enclosed in the model, highlight the significant target features, at the same time reduce the influence of the features shared by both the target and the background, and exhibit an excellent monotonic property and a distinct peak-like distribution [19].

The aim of image matching is to search locally for a sub-area within an image which is most similar to a predefined target model. The sub-areas to be matched against the target model is called target candidate. To reduce the computational cost, m_u -bin histograms are used. Thus, we have the target model $q = \{q_u\}_{u=1, \dots, m_u}$ and the target candidate $p = \{p_u\}_{u=1, \dots, m_u}$. Assuming s is the un-normalized histogram vector of the search region, PPM which is an evaluation of the posterior probability of a target candidate to be identified as the target can be defined as

$$\phi(p, q) = \frac{1}{m} \sum_{u=1}^{m_u} \frac{p_u q_u}{s_u}. \quad (3)$$

where s_u , p_u , and q_u represent the u th elements of the un-normalized histogram vectors s , p , and q , respectively; m is the number of pixels in the target model, which is a normalization constant.

Equation (3) can be transformed to the pixel-wise form:

$$\phi(p, q) = \frac{1}{m} \sum_{j=1}^m \frac{q_u(j)}{s_u(j)}. \quad (4)$$

where j is the pixel index, $q_u(j)$ and $s_u(j)$ represent the u th bin value of the target model histogram and the search region histogram, respectively. Obviously, we can regard the value of $q_u(j)/s_u(j)$ as the contribution of pixel j to the similarity value. Thus, the PPM can be computed by accumulating individual pixel's contributions.

The pixel-wise PPM in (4) creates conditions for scale adaptation because it is simply a linear combination of contributions from individual pixels. So the PPM can be averaged to obtain an approximate contribution of each pixel to the similarity measure as

$$\bar{\phi} = \frac{\phi(p(y^*), q)}{m}. \quad (5)$$

where y^* denotes the current target center position. The average similarity contributions of the pixels in the inner and outer layers are denoted as

$$\bar{\phi}_i, i = -a, \dots, 0, \dots, a \quad (6)$$

where $i < 0$ means the i th inner layer, $i > 0$ means the i th outer layer, and $i = 0$ represents the target region border.

An empirical scale adaptation formula is given as

$$w(k+1) = \begin{cases} w(k)+2 & \text{if } \bar{\phi}_{-1} > 0.8 \text{ and } \bar{\phi}_0 > 0.75 \text{ and} \\ & \bar{\phi}_1 > 0.6 \\ w(k)-2 & \text{if } \bar{\phi}_0 < 0.6 \text{ and } \bar{\phi}_1 < 0.3 \\ w(k) & \text{else} \end{cases} \quad (7)$$

where $w(k)$ is the size of the target region at frame k .

The detailed object localization and scale adaptation process with PPM was introduced in [28].

III. EVOLUTION OF OBJECT CONTOUR

After the target localization with the PPM, the object feature image is generated after a series of pre-processing such as background subtraction or using the color probability density information in the object region. Then we adopt the momentum based level set segmentation to evolve the object contour in the object feature space, in order to improve the tracking precision.

A. The Momentum Based Level Set Segmentation

Level set method was proposed by Osher and Sethian, which can compatibly solve the problem of arbitrary topology change. To improve the convergence and the search speed, while reaching the global optima as fast as possible, in this paper, we use a momentum based level set segmentation method by simulating the physical properties of inertia and momentum and effectively allowing the search to avoid local optima and accelerating in favorable directions [29]. It has employed a simple energy functional based on a weighted region term combined with a penalty on curve length for regularization. Obviously, our goal is to maximize:

$$E(C) = \iint_{\Omega_c} f(x, y) dx dy - \alpha \int_C ds \quad (8)$$

where C is a 1D curve embedded in a 2D domain, Ω_c is the inside region of C , $f(x, y)$ is a scalar function called cost function and α is a regularization weight parameter.

Translating (8) to a level set partial differential equation (PDE) so we acquire a familiar level set equation:

$$\frac{\partial \phi}{\partial t} = -f(x, y) |\nabla \phi| + \alpha \kappa |\nabla \phi| \quad (9)$$

where κ is the curvature of the contour.

Turning to gradient descent with momentum, a search direction is chosen according to:

$$s_k = -\eta(1-\omega)\nabla f_k + \omega s_{k-1} \quad (10)$$

where η is the learning rate and $\omega \in [0, 1]$ represents the momentum.

The difference between two subsequent time instances of the level set function is taken to approximate the direction of the gradient, and the entire level set function can be represented as one vector:

$$\nabla f(t_n) \approx \frac{\phi(t_n) - \phi(t_{n-1})}{\Delta t} \quad (11)$$

Note that this is really an approximation, depending on the time interval between two adjacent points $\Delta t = t_n - t_{n-1}$. Therefore we update the level set function with a momentum term:

$$s(t_n) = -\eta(1-\omega) \frac{\hat{\phi}(t_n) - \phi(t_{n-1})}{\Delta t} + \omega s(t_{n-1}) \quad (12)$$

$$\phi(t_n) = \phi(t_{n-1}) + \Delta t s(t_n) \quad (13)$$

The procedure of Update Level set is:

- Step1: Given the level set function $\phi(t_{n-1})$, the next intermediate time step $\hat{\phi}(t_n)$ can be computed according to (9).
- Step2: Compute the approximate gradient by (11).
- Step3: Compute a step $s(t_n)$ according to (12).

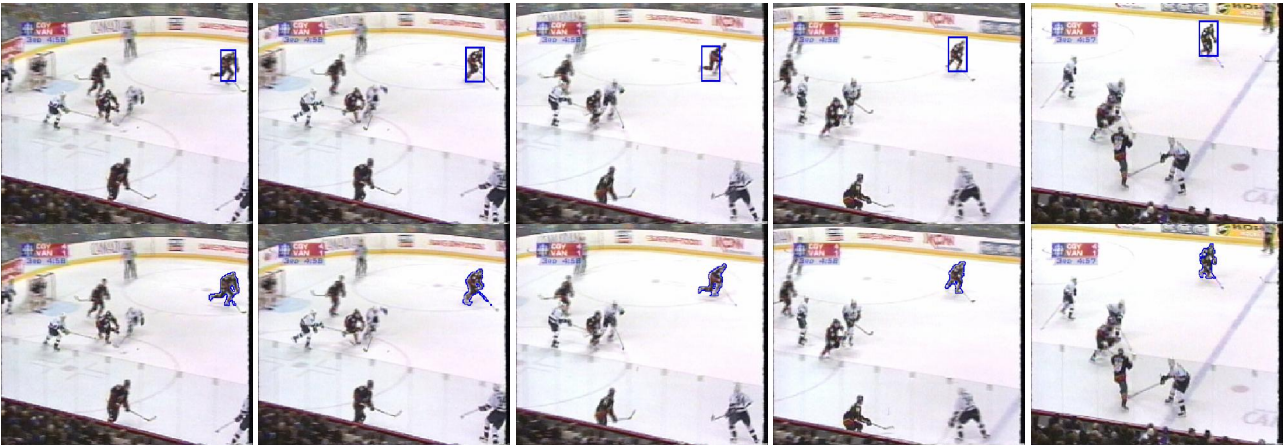


Figure 1. Tracking of the ice hockey sequence with the kernel-based method (top) and our method (bottom). Frames 1, 6, 13, 21, and 37 are shown.



Figure 2. Tracking of the vehicle1 sequence with the kernel-based method (top) and our method (bottom). Frames 1, 9, 19, and 23 are shown.

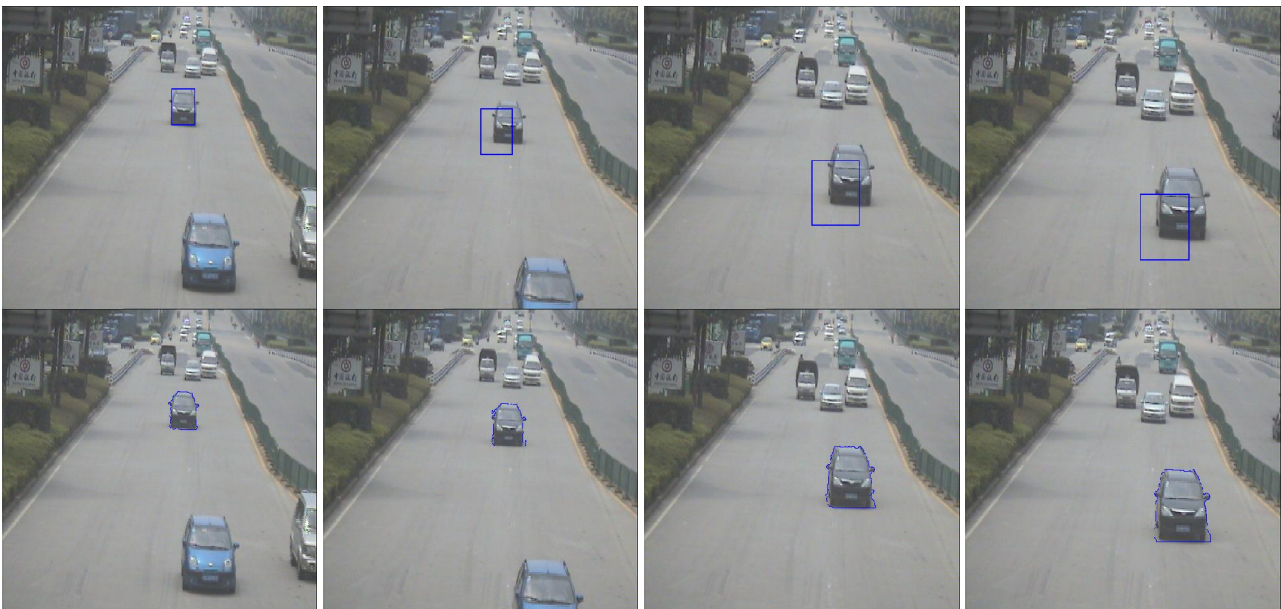


Figure 3. Tracking of the vehicle2 sequence with the kernel-based method (top) and our method (bottom). Frames 1, 10, 30, and 37 are shown.

- Step4: Compute the next time step $\phi(t_n)$ by (13) and the intermediate level set function computed in Step 1 will be replaced.

B. Evolution Control

We define the convergence condition as $|\nabla f|_\infty < 0.03$, with ∇f given in (13). If the convergence condition is satisfied, iteration stops, otherwise, iteration continues.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to test and verify the effectiveness of proposed algorithm, we did some experiments of videos, and have implemented our idea with Matlab7.0.1 on

Windows Xp, AMD Athlon(tm) 64 X2 Dual Core CPU, 1.9GHz, 2GB RAM.

Fig. 1 shows the tracking results of the ice hockey sequence with the kernel-based method [15] and our method, in which the frames with deformation are shown. The ice hockey sequence has 40 frames of 240×320 pixels. The scale of the kernel is adjusted with $\pm 10\%$ of the current size. The color spaces for our method and the kernel-based method are both quantized into $16 \times 16 \times 16$ bins. Fig. 2 and Fig. 3 show the tracking results of two vehicle sequences that obtained by ourselves with the kernel-based method [15] and our method, which has 28 frames of 576×704 pixels and 37 frames of 576×704 pixels, respectively. The first frames are all the manual initialization, in which we can see that a higher

precision target template is obtained with contour segmentation algorithm, eliminating the background interference in a way. In Fig. 1, because of background pixels included in target template and deformation of the athlete while fast moving, a bias occurs in frame 13. As for two vehicle sequences, since the traffic system usually track the vehicles for some time to get useful information like velocity, licensing information and so on, we just select a certain drive distance to track. The environment of the road becomes more complex than the ice hockey sequence, so the difficulty of the tracking task is higher. In Fig. 2, the kernel-based method has a bias through the sequence and the bias becomes more obvious when a motorcycle appears from the back of the target, while our method performs very well. In Fig. 3, the kernel-based method maintains a distinct bias owing to the similarity of the colour between the road and the target.

In order to evaluate the tracking precision quantitatively, we adopt the error measure as follows:

$$error = 1 - \frac{(O_1 \cap O_2) \cup (B_1 \cap B_2)}{O_1 \cup B_1} \quad (14)$$

where O_1 and B_1 are the object region and background region in the manual segmentation result, O_2 and B_2 are the object region and background region in the segmentation result with the object tracking method. Meanwhile, the location error of object center is adopted to evaluate the location precision, which is defined as follows:

$$d = \|C_a - C_m\| \quad (15)$$

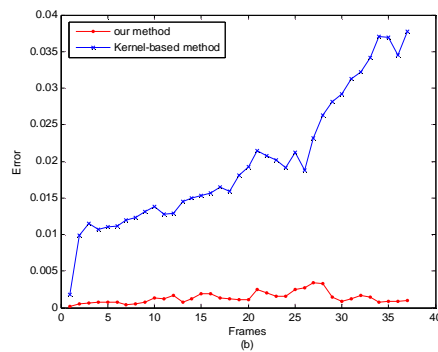
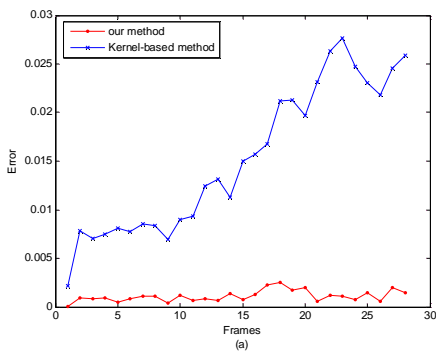


Figure 4. Comparison of the tracking error. (a) Tracking error of Fig. 2. (b) Tracking error of Fig. 3.

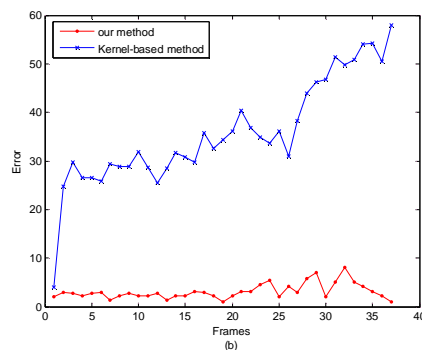
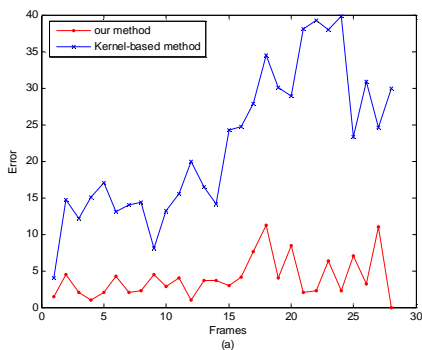


Figure 5. Comparison of the location error. (a) Location error of Fig. 2. (b) Location error of Fig. 3.

where C_a and C_m are the object center coordinates of the object tracking and manual segmentation results, respectively. Fig. 4 which shows the tracking errors of Figs. 2 and 3, respectively, indicates that our method can cope with the deformation and the scale change more effectively than the kernel-based method, so our method can achieve a higher tracking precision. Fig. 5 which shows the location errors of Figs. 2 and 3, respectively, indicates that our method can reduce the influence of the background and has the better location of object center than the kernel-based method for most frames. The average time performance of our method in the ice hockey sequence is about 0.16s per frame, it can reach the requirement of real-time tracking. But when the target becomes bigger like vehicles, the average time per frame gets longer which is about 3s.

V. CONCLUSION

By combining the merits of the region-based method and the contour-based method, we have presented a robust object tracking method. Using the Kalman filter and PPM, we can locate the object effectively in complex condition with partial occlusions, camera motion, clutter, object deformation etc. In order to improve the tracking precision, we used the contour-based method to track the object contour precisely after the target localization. The experimental results demonstrated that our method can achieve a higher tracking precision than that of the kernel-based method, but our method is not competent for real-time tracking when the target is large and fast moving.

In the future, an important direction for us to study is

that how to combine PPM with other fast algorithms such as mean-shift to improve the time performance and how to remove the shadow entirely. What's more, we will further study to propose more efficient methods to extract the contour also for saving time.

ACKNOWLEDGMENT

We would like to thank Zhejin Wang for his helpful discussions. We would also like to thank Peidong Zhou for dealing with the problems of typesetting. This work was supported by the National Natural Science Foundation of PRC (No.50908213).

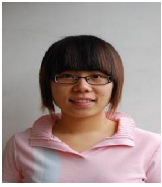
REFERENCES

- [1] M. Greiffenhagen, D. Comaniciu, H. Niemann and V. Ramesh, "Design, Analysis and Engineering of Video Monitoring Systems: An Approach and a Case Study", Proc. IEEE, vol. 89, no. 10, pp.1498-1517, 2001.
- [2] G. Bradski, "Computer Vision Face Tracking as a Component in a Perceptual User Interface", Proc. IEEE Workshop Applications of Computer Vision, pp. 214-219, 1998.
- [3] A. Eleftheriadis and A. Jacquin, "Automatic Face Location Detection and Tracking for Model-Assisted Coding of Video Teleconferences sequences at low bit-rate", Signal Processing: Image Communication, vol. 7, no. 4-6, pp. 231-248, 1995.
- [4] U. Handmann, T. Kalinke, C. Tzomakas, M. Werner and W. Seelen, "Computer Vision for Driver Assistance Systems", Proc. SPIE, vol. 3364, pp. 136-147, 1998.
- [5] E.S. Baker, R.D. DeGroat, "A correlation-based subspace tracking algorithm", IEEE Trans. Signal Process, vol. 46, no. 11, pp. 3112-3116, 1998.
- [6] L.G. Brown, "A survey of image registration techniques", ACM Comput. Surveys, vol. 24, no. 4, pp. 325-376, 1992.
- [7] B.K.P. Horn, B.G. Schunk, "Determining optical flow", Artif. Intell, vol. 17, no. 13, pp. 185-203, 1981.
- [8] J. Malki, L. Mascarilla, E.H. Zahzah, P. Boursier, "Directional relations composition by orientation histogram fusion", IEEE Conference on Pattern Recognition, vol. 3, pp. 758-761, 2000.
- [9] S.M. Smith, J.M. Brady, "real-time motion segmentation and shape tracking", IEEE Trans. Pattern Anal. Mach. Intell, vol. 17, no. 8, pp. 814-820, 1995.
- [10] B. Zitova, J. Flusser, "Image registration methods: a survey", Image Vis. Comput, vol. 21, pp. 977-1000, 2003.
- [11] A. Djouadi, O. Snorrason, F. Garber, "The quality of training-sample estimates of the Bhattacharyya coefficient", IEEE Trans. Pattern Anal. Mach. Intell, vol. 12, pp. 92-97, 1990.
- [12] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection", IEEE Trans. Commun. Technol, vol. 15, pp. 52-60, 1967.
- [13] J. Lin, "Divergence measures based on the Shannon entropy", IEEE Trans. Inf. Theory, vol. 37, pp. 145-151, 1991.
- [14] T.L. Liu, H.T. Chen, "Real-time tracking using trust-region methods", IEEE Trans. Pattern Anal. Mach. Intell, vol. 26, no. 3, pp. 397-402, 2004.
- [15] D. Comaniciu, V. Ramesh, P. Meer, "Kernel-based object tracking", IEEE Trans. Pattern Anal. Mach. Intell, vol. 25, no. 5, pp. 564-577, 2003.
- [16] D. Comaniciu, V. Ramesh, P. Meer, "Real-time tracking of non-rigid objects using mean shift", IEEE Comput. Vision Pattern Recognition, vol. 2, pp. 142-149, 2000.
- [17] D. Comaniciu, P. Meer, "Mean shift: a robust approach toward feature space analysis", IEEE Trans. Pattern Anal. Mach. Intell, vol. 24, no. 5, pp. 603-619, 2002.
- [18] Y. Cheng, "Mean shift, mode seeking, and clustering", IEEE Trans. Pattern Anal. Mach. Intell, vol. 17, no. 8, pp. 790-799, 1995.
- [19] G. Hager, M. Dewan, C. Stewart, "Multiple kernel tracking with ssd", IEEE Conference on Computer Vision Pattern Recognition, vol. 1, pp. 790-797, 2004.
- [20] C. Yang, R. Duraiswami and L. Davis, "Efficient mean-shift tracking via a new similarity measure", Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit. (CVPR), vol. 1, pp. 176-183, 2005.
- [21] D. Comaniciu, "An algorithm for data-driven bandwidth selection", IEEE Trans Pattern Anal. Mach. Intell, vol. 25, no. 2, pp. 281-288, 2003.
- [22] D. Comaniciu, V. Ramesh, P. Meer, "The variable bandwidth mean shift and data-driven scale selection", IEEE Int. Conf. Comput. Vision, vol. 1, pp. 438-445, 2001.
- [23] R.T. Collins, "Mean-shift blob tracking through scale space", IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition, vol. 2, pp. II-234-40, 2003.
- [24] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras", IEEE Trans. Pattern Anal. Mach. Intell, vol. 26, no. 11, pp. 1531-1536, 2004.
- [25] C.Y. Chung and H.H. Chen, "Video object extraction via MRF-based contour tracking," IEEE Trans. Circuits Syst. Video Technol, vol. 20, no. 1, pp. 149-155, 2010.
- [26] R.E. Kalman, "A new approach to linear filtering and prediction problems," Trans. Am. Soc. Mechan. Eng.-J. Basic Eng, vol. 82, no. 1, pp. 35-45, 1960.
- [27] G. Welch and G. Bishop, "An introduction to the Kalman filter," Dept. Comput. Sci, pp. 95-041, 2004.
- [28] Z.R. Feng, N. Lu, and P. Jiang, "Posterior probability measure for image matching", Pattern Recognition, vol. 41, pp. 2422 - 2433, 2008.
- [29] G. L  th  n, T. Andersson, R. Lenz, and M. Borga. "Momentum Based Optimization Methods for Level Set Segmentation", SSVM 2009, LNCS 5567, pp. 124-136, 2009.



etc.

Haocheng. Le received his B. Sc. degree in automation from Zhejiang University of Technology, China, in 2010, and he is currently a graduate student in Zhejiang University of Technology, China. His research interests include computer vision and pattern recognition, artificial intelligence,



Linglong. Hu will receive the bachelor's degree in automation engineering from Zhejiang University of Technology, China, in 2011. And she has been recommended to Zhejiang University in Control Science and Engineering for her graduate study. Her research interests include ramp metering and

route recommendation in intelligent traffic field.



Yuanjing. Feng received his B. Sc. in Mechanical Engineering and M. Sc. in Mechanical Engineering from the Northwest A&F University, Xi'an, China, in 1998 and 2001, respectively, and Ph. D. degree in Control Science and Engineering from Xi'an Jiaotong University, Xi'an, China in 2005. He

is an associate professor in the School of Information Engineering at Zhejiang University of Technology, PRC from 2005 on. His research interests include image analysis and understanding, etc.