Feature Particles Tracking for Moving Objects

Tao Gao
School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China
Email: gaotao231@yahoo.cn

Ping Wang
School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China
Email: wangps@tju.edu.cn

Chengshan Wang
School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China
Email: cswang@tju.edu.cn

Zhenjing Yao
Department of Disaster Prevention Instrument, Institute of Disaster Prevention Science and Technology, Beijing 101601, China
Email: yaozhenjing@cidp.edu.cn

***Abstract*—For particle filtering tracking method, particle choosing was random to some degree according to the dynamics equation, which may cause inaccurate tracking results. To compensate, an improved particle filtering tracking method was presented. A moving vehicle was detected by redundant discrete wavelet transforms method (RDWT), and then the key points were obtained by scale invariant feature transform. The matching key points in the follow-up frames obtained by SIFT method were used as the initial particles to improve the tracking performance. Experimental results show that more particles centralize in the region of motion area by the presented method than traditional particle filtering, and tracking results of moving vehicles are more accurate; the run-time is only 0.15s more than traditional method, indicating that it has a certain actual application prospect.**

***Index Terms*—Moving vehicle tracking, Particle filtering, SIFT, RDWT, ITS**

I. INTRODUCTION

Video surveillance system is an important part of Intelligent Transportation System (ITS). It is the applications of new information and communication technologies (ICT) into vehicles and roadways for monitoring traffic conditions, reducing congestion, enhancing mobility, optimizing the use of transport infrastructure and improving security. They also draw traffic forecasts and advise the best departure times in different regions, and alternative routes for major roads. They inform road users in real time on their travel time and best routes to be taken given a destination, and so on. The technologies used in intelligent transportation systems range from basic management systems such as systems management signalized intersections, systems management containers, variable message signs (VMS), automatic radar or video surveillance to applications more advanced that integrate data in real time with feedback from many sources, such as weather information, bridges de-icing systems, embedded navigation systems informing travel time in realtime. Additionally, predictive techniques have been developed to allow advanced modeling and comparison with a database of historical references.

Detecting and tracking moving vehicles from video sequences is one of the important tasks in video surveillance system. Recently, many approaches have been proposed in this field. Ref.[1] use a template matching method to track the target. The blobs correspond to the moving target in the video sequences. But the method is difficult in handling scale change of the target, and threshold is subjectively determined with less robustness. Ref.[2] uses a snake model based tracking method which can reduce computational complexity and improve tracking accuracy. But it is sensitive to initialization and is difficult for actual application. Ref.[3] present a mean-shift method for motion tracking. Mean-shift method manifests high efficiency for target tracking with low complexity. But as a hill climbing algorithm, it may fall into a local minimum and lose the motion target when occlusion occurs. Ref.[4] uses Particle filtering to track a moving target; it is a successful numerical approximation technique for Bayesian sequential estimation with non-linear, non-Gaussian models. The basic Bayesian filtering is a recursive process in which each iteration consists of a prediction step and a filtering step. There are also other many methods in the literature, such as thresholding, multi-resolution processing, edge

Project number: 71102174, 12213519D and 10ZCKFSF01100.
Contact author: Tao Gao

doi:10.4304/jmm.7.6.408-414

detection, background subtraction and inter-frame differencing [5] [6]. Background subtraction is the more accurate method for traffic monitoring and it is widely exploited in many other applications, such as human motion capture, and video surveillance. This is a technique to remove non moving components from a video sequence. The main assumption for its application is that the camera remains stationary. The basic principle is to create a reference frame of the stationary components in the image and then each pixel of the prototype is compared with the actual image color map. If the color difference exceeds a predefined threshold it is assumed that this pixel belongs to the foreground. This algorithm is carried out in three steps: Initialization of the background, Foreground extraction, Background update.

In this paper, the positions of particles are determined by key points obtained by scale invariant feature transform (SIFT) to improve the tracking efficiency, and we organize the paper as follows. A brief introduction on RDWT motion detection is given in Section 2. Scale invariant feature transform method is described in Section 3. The improved particle filtering tracking method [7] [8] is described in Section 4. Experimental results are reported in Section 5. Finally conclusions are summarized in Section 6.

II. Moving Vehicle Detection

In this paper, moving vehicles are detected by redundant discrete wavelet transforms (RDWT) [9], which conquer the drawback of time-domain methods. The RDWT is an approximation to the continuous wavelet transform that removes the down-sampling operation from the traditional critically sampled DWT to produce an over-complete representation. The shift-variance characteristic of the DWT arises from its use of down-sampling; while the RDWT is shift invariant since the spatial sampling rate is fixed across scale. As a result, the size of each sub-band in an RDWT is the exactly the same as that of the input signal. The redundant wavelet transforms for an image $p(i, j)$ can be in such ways:

$$PLL_j(x,y) = PLL_{j-1}(x,y) * ([h]_{\uparrow 2^{j-1}}, [h]_{\uparrow 2^{j-1}})(-x,-y);$$

$$PLH_j(x,y) = PLH_{j-1}(x,y) * ([h]_{\uparrow 2^{j-1}}, [g]_{\uparrow 2^{j-1}})(-x,-y);$$

$$PHL_j(x,y) = PHL_{j-1}(x,y) * ([g]_{\uparrow 2^{j-1}}, [h]_{\uparrow 2^{j-1}})(-x,-y);$$

$$PHH_j(x,y) = PHH_{j-1}(x,y) * ([g]_{\uparrow 2^{j-1}}, [g]_{\uparrow 2^{j-1}})(-x,-y).$$

Because the coefficients of the sub-bands of the redundant wavelet transform are highly correlated, and the direction and size are the same as the image, also, there is no translation in the sub-band; this paper uses the method which is based on redundant wavelet transforms to obtain the motion area. First, if the two adjacent frames are f_1 and f_2, we use the equation (1) to obtain the $MAS(x,y)$:

$$MAS(x,y) = \sum_{j=J_0}^{J_1} \left(\begin{array}{l} \left| LL_1^{(j)}(x,y) - LL_2^{(j)}(x,y) \right| + \left| LH_1^{(j)}(x,y) - LH_2^{(j)}(x,y) \right| \\ + \left| HL_1^{(j)}(x,y) - HL_2^{(j)}(x,y) \right| + \left| HH_1^{(j)}(x,y) - HH_2^{(j)}(x,y) \right| \end{array} \right) \quad (1)$$

J_0 and J_1 are the starting and ending scales. We can obtain the motion area according to equation (2):

$$motion(x,y) = \begin{cases} 1 & MAS(x,y) \geq T \\ 0 & MAS(x,y) < T \end{cases}. \quad (2)$$

T is the threshold which can be obtained automatically by Otsu [10] method, then the mathematical morphology is used to remove noise points. The binary motion mask obtained from the redundant wavelet transforms can be considered as the original mask of the moving object. Moreover, shadow points are usually adjacent to object points and with the more commonly used segmentation techniques shadows and objects are merged in a single blob. This affects both the classification and the assessment of moving object position. As the inner district of the object is usually flat and the characteristic is not obvious, this paper uses an assimilation method [11] to fill the mask.

III. Sift Key Points

Feature extraction and vehicles classification are used to extract real time traffic information, such as the vehicle count, traffic events, and traffic flow, which play an important role for traffic analysis and traffic management. There are many different approaches based on image processing. Some works use entropy as an underlying measurement to calculate traffic flows and vehicle speeds, others used bounding box and Kalman filter to track each detected vehicle. Moreover, 3D vehicle features and models, and several vehicle features like shape, length, width, height, texture, etc. It can be extracted to classify vehicle into different categories and to determine traffic parameters.

So after getting the vehicle region, we use scale invariant feature transform (SIFT) to obtain the key points in the vehicle region. As SIFT transforms image data into scale-invariant coordinates relative to local features, an important aspect of this approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations [12] [13]. However a given object can appear differently depending on three parameters: its distance from the camera, its eccentricity from the axis of the eye and its incline in three directions. Therefore, to minimize these affects, before recognition, a process of camera calibration for normalizing features should be applied. For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors. There are four steps for this method.

1) Find Scale-Space Extrema. The only reasonable kernel for scale-space (continuous function of scale σ) is Gaussian:

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2))/2\sigma^2}\,. \qquad (3)$$

For two-dimensional image, $L(x,y,\sigma)=G(x,y,\sigma)*I(x,y)$. Experimentally, Maxima of Laplacian-of -Gaussian: $\sigma^2\nabla^2 G$ gives best notion of scale. As LOG is expensive, we define Difference-of- Gaussians (DoG) instead:

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma)\,. \qquad (4)$$

The smoothed images need to be computed in any case for feature description, but in application we need only to subtract two images, and then choose all extrema within 3x3x3 neighborhood.

2) Remove the Instable and Edge Points. Take Taylor series expansion:

$$D\left(\vec{x}\right)=D+\frac{\partial D^T}{\partial\vec{x}}\vec{x}+\frac{1}{2}\vec{x}^T\frac{\partial^2 D^T}{\partial\vec{x}^2}\vec{x}\,. \qquad (5)$$

Minimize it to get true location of extrema: $\hat{X}=-\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X}$, where $D(\hat{X})=D+\frac{1}{2}\frac{\partial D^{-1}}{DX}\hat{X}$. We remove the instable points by the rule: $\left|D(\hat{X})\right|<0.03$, and also reject points which do not satisfy the equation (6):

$$\frac{(Tr(H))^2}{Det(H)}<\frac{(r+1)^2}{r} \qquad (6)$$

Where $Tr(H)=D_{xx}+D_{yy}$, $Det(H)=D_{xx}D_{yy}-(D_{xy})^2$, r =10.

3) Orientation Assignment. We use scale of point to choose correct image:

$$L\left(x,y\right)=G\left(x,y,\sigma\right)*I\left(x,y\right) \qquad (7)$$

Compute gradient magnitude and orientation using finite differences:

$$m\left(x,y\right)=\sqrt{\left(L\left(x+1,y\right)-L\left(x-1,y\right)\right)^2+\left(L\left(x,y+1\right)-L(x,y-1)\right)^2} \qquad (8)$$

$$\theta\left(x,y\right)=\tan^{-1}\left(\frac{\left(L\left(x,y+1\right)-L(x,y-1)\right)}{\left(L\left(x+1,y\right)-L\left(x-1,y\right)\right)}\right) \qquad (9)$$

4) Obtain the Matching Points. The error of SIFT descriptors of the extrema points of image 1 and 2: $error=(\widehat{F}1-\widehat{F}2)^2$ can be used to obtain the matching points. Figure 1 show the key points with orientation and SIFT matching result of a vehicle.

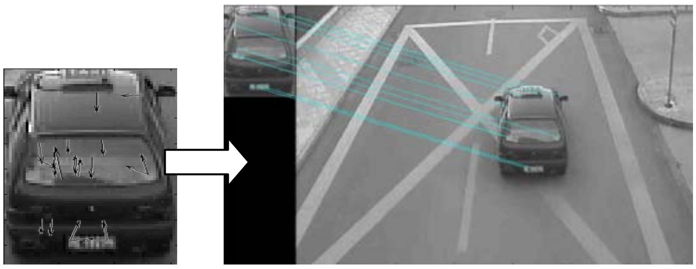


Figure 1. Key points with orientation and matching.

IV. Active Particle Filtering Combined With SIFT

Particle filtering [14] [15] essentially combines the particles at a particular position into a single particle, giving that particle a weight to reflect the number of particles that were combined to form it. This eliminates the need to perform redundant computations without skewing the probability distribution. Particle filtering accomplishes this by sampling the system to create N particles, then comparing the samples with each other to generate an importance weight. After normalizing the weights, it resamples N particles from the system using these weights. This process greatly reduces the number of particles that must be sampled, making the system much less computationally intensive. Particle Filtering estimates the state of the system, x and t, as time t as the Posterior distribution: $P(x_t \mid y_{0-t})$. Let $Es(t) = P(x_t \mid y_{0-t})$, $Es(1)$ can be initialized using prior knowledge Particle filtering assuming a Markov Model for system state estimation. Markov model states that past and future states are conditionally independent given current state. Thus, using Markov model, observations are dependent only on current state:

$$Es(t) = P(x_t \mid y_{0-t}) = P(y_t \mid x_t, y_{0-t-1})P(x_t \mid y_{0-t-1}) = P(y_t \mid x_t)P(x_t \mid x_{t-1})P(x_{t-1} \mid y_{0-t-1}) = P(y_t \mid x_t)P(x_t \mid x_{t-1})Es(t-1) \quad \textbf{(10)}$$

Final Result:

$$Es(t) = P(y_t \mid x_t)P(x_t \mid x_{t-1})Es(t-1) \quad \textbf{(11)}$$

Where: $P(y_t \mid x_t)$ is observation model and $P(x_t \mid x_{t-1})Es(t-1)$ is proposal distribution. The basic model usually consists of a Markov chain X and a possibly nonlinear observation Y with observational noise V independent of the signal X . System Dynamics Motion Model is $P(x_t \mid x_{0:t-1})$, and Observation Model is $P(y_t \mid x_t)$, Posterior Distribution is $P(x_t \mid y_{0 \cdots t})$. Proposal Distribution is the Motion Model Weight, W_t = Posterior / Proposal = observation. Given N particles $\{x^{(i)}{}_{0:t-1}, z^{(i)}{}_{0:t-1}\}^{N}{}_{i=1}$ at time $t-1$, approximately distributed according to the distribution: $P(dx^{(i)}_{0:t-1}, z^{(i)}_{0:t-1} \mid y_{1:t-1})$, particle filters enable us to compute N particles $\{x^{(i)}{}_{0:t}, z^{(i)}{}_{0:t}\}^{N}{}_{i=1}$ approximately distributed according to the posterior distribution $P(dx^{(i)}_{0:t}, z^{(i)}_{0:t} \mid y_{1:t})$. In video tracking, we do as follows: for each particle at time t, we sample from the transition beforehand. For each particle, we evaluate and normalize the importance weights, then multiply or discard particles with respect to high or low importance weights $W_t^{(i)}$ to obtain N particles. This selection step is what allows us to track moving objects efficiently. The state space is represented in the spatial domain as: $X = (x, y)$. We have initialized the state space for the first frame automatically by using the matching key points obtained by SIFT. A second-order auto-regressive dynamics is chosen on the parameters by SIFT matching to represent the state space (x, y) . The dynamics is given as: $X_{t+1} = Ax_t + Bx_{t-1}$. Matrices A and B could be learned from a set of sequences where correct tracks have been obtained. In this paper, $A = \begin{bmatrix} 1 & \sigma \\ 0 & 1 \end{bmatrix}$, and $B = \alpha \begin{bmatrix} \frac{\sigma^3}{3} & \frac{\sigma^2}{2} \\ \frac{\sigma^2}{2} & \sigma \end{bmatrix}$, where $\sigma = 3$, and $\alpha = 0.35$.

The observation y_t is proportional to the histogram distance between the color window of the predicted location in the frame and the reference color window: $Dist(q, q_X)$, Where q = reference color histogram, q_X = color histogram of predicted location. The following pseudo code depicts the overall structure of our tracking system.

At time $t+1$, construct the n^{th} of N samples as follows:

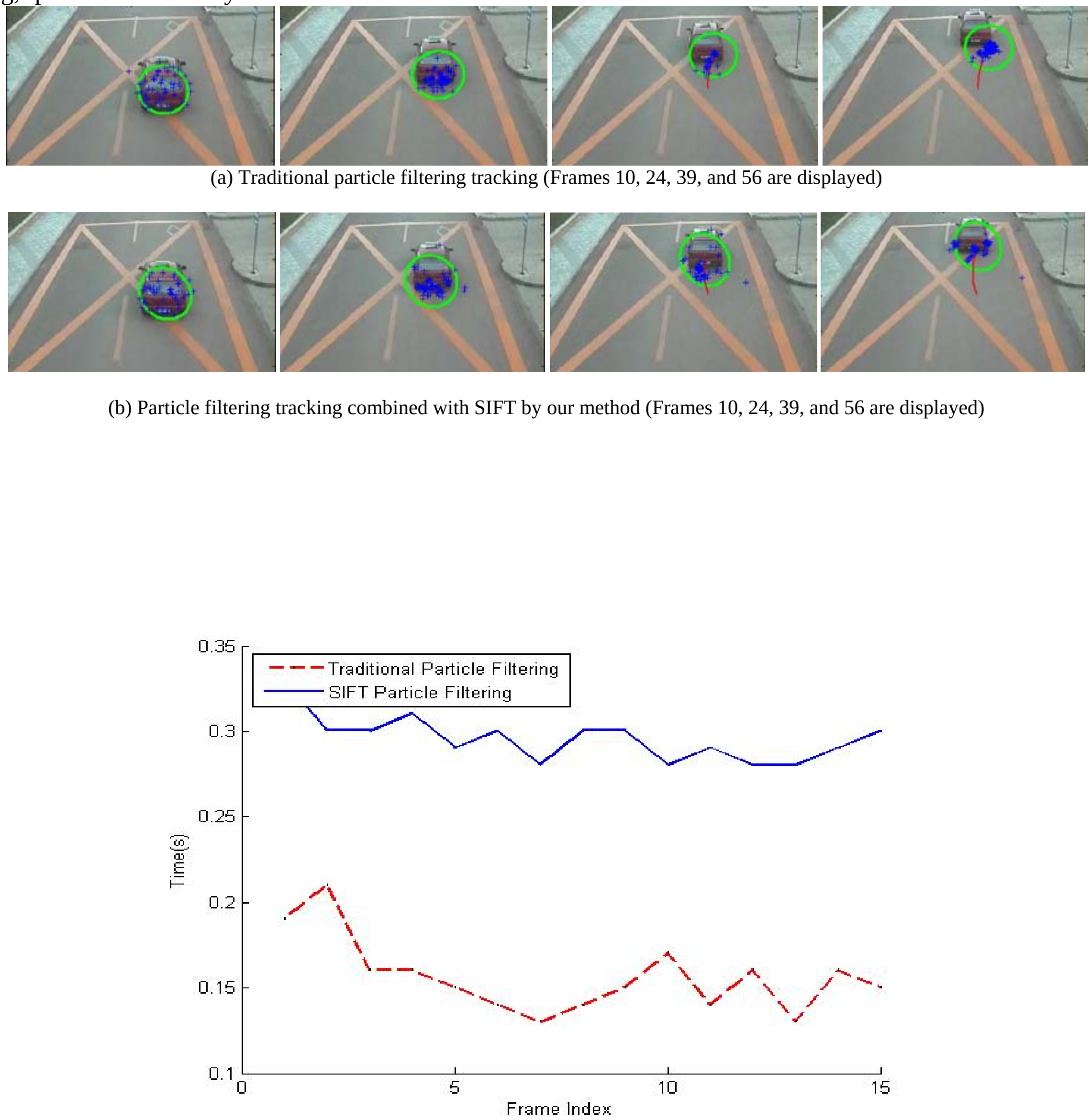
```
1. Generate a random number $r \in [0,1]$, uniformly
   distributed.
2. Find, by binary subdivision on m, the smallest m
   for which $c_t^{(m)} \geq r$ .
3. First, draw a random variate $s_{t+1}^{(n)}$ from the
   density $p(x_{t+1} \mid x_t = s_t^{(m)})$, and then the first
   $k$ values of $s_{t+1}$ are set by matching key points
   of $SIFT(t, t+1)$.
4. Store samples $n = 1, \cdots, N$ as
   $(s_{t+1}^{(n)}, \pi_{t+1}^{(n)}, c_{t+1}^{(n)})$ where
       $c_{t+1}^{(n)} = 0$
       $\pi_{t+1}^{(n)} = p(z_{t+1} \mid x_{t+1} = s_{t+1}^{(n)})$
       $c_{t+1}^{(n)} = c_{t+1}^{(n-1)} + \pi_{t+1}^{(n)}$
   and then normalize by dividing all cumulative
   probabilities $c_{t+1}^{(n)} = c_{t+1}^{(N)}$, that $c_{t+1}^{(N)} = 1$.
```

Mean properties can be estimated at any time t as:

$$E[f(x) \mid z_t] \approx \sum_{n=1}^{N} \pi_t^{(n)} f(s_t^{(n)}).$$

V. Experimental Results

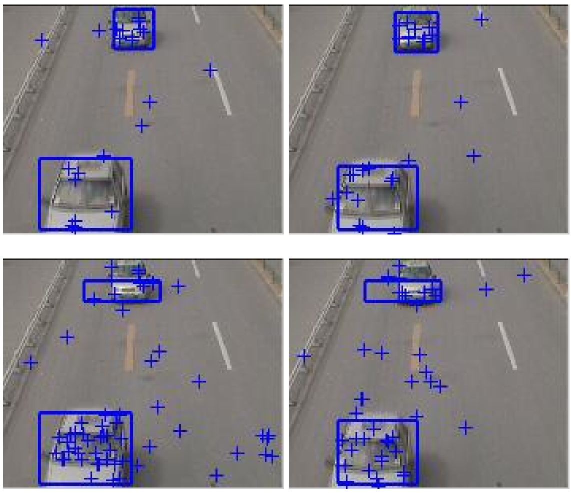
In traditional particle filter tracking method, particle choosing is random according to the dynamics equation in some degree, which may cause inaccurate tracking results. In this section, we compare the tracking results between traditional particles filtering and our method showed in figure 2 (a) and (b). The video is sampled at a resolution of 768x576 and a rate of 25 frames per second. The algorithms are tested on a 1400 MHz Celeron CPU, and software environment is VC++ 6.0. From the results we can see that when the scale of vehicle changes drastically, the particles still locate in the region of vehicle by our method; while for traditional particle filtering, particles obviously deviate from the vehicle which causes inaccurate results. The blue cross sign shows the particle, and red curve shows the motion track. We also sample 15 frames from the video sequence and compare the runtime of traditional particle filtering and our method (SIFT Particle Filtering), also showed in figure 2 (c); the difference is about 0.15s which can be ignored in actual application.



(a) Traditional particle filtering tracking (Frames 10, 24, 39, and 56 are displayed)

(b) Particle filtering tracking combined with SIFT by our method (Frames 10, 24, 39, and 56 are displayed)

(c) Runtime comparison



(d) Comparison of multiple targets tracking

Figure 2. Comparisons of performance.

Occlusion is an overlapping of objects from the viewpoint of a camera. In this case, multiple vehicles in the occlusion are identified as one mobile object (vehicle). The two major factors to cause occlusions are perspective effects and shadows. The extracted mobile objects from background subtraction often include the shadows. For multiple targets tracking and shadows influence, each target can be tracked with the same process separately. Compared with ref. [4], figure 2 (d) shows the better accuracy by our method. First row shows tracking results by our method, and second row shows ref. [4] tracking results. Experimental results show that the proposed methods offer great improvements in terms of accuracy, robustness, and stability in traffic surveillance.

VI. Conclusions

In this paper, a novel moving vehicle tracking method based on particle filtering and SIFT is presented. First, the motion vehicle is detected by redundant discrete wavelet transforms, and scale invariant feature transform is used to extract the key points of vehicle; then according to SIFT matching, initial positions of particles can be obtained. By actively choosing the particles, tracking performance can be significantly improved.

Acknowledgment

This work is supported by National Natural Science Foundation (NNSF) of China (No. 71102174), the Science Foundations of Tianjin under Grant No. 10ZCKFSF01100, and the Hebei Science and Technology Support Project under Grant No. 12213519D.

References

[1] D. R. Magee, “Tracking Multiple Vehicles Using Foreground, Background and Motion Models”, *Image and Vision Computing*, vol. 22, no. 2, pp. 143–155, 2004.

[2] Haoting Liu, Guohua Jiang, Li Wang, “Multiple Objects Tracking Based on Snake Model and Selective Attention Mechanism”, *International Journal of Information Technology*, vol. 12, no. 2, pp.76-86, 2006.

[3] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based Object Tracking”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564 - 577, 2003.

[4] C. Hue, J. Le Cadre, and P. Perez, “Tracking Multiple Objects with Particle Filtering”, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38 , no. 3, pp. 791 - 812, 2003.

[5] H. Dahlkamp, H.-H. Nagel, A. Ottlik, and P. Reuter, “A framework for model-based tracking experiments in image sequences”, *International Journal of Computer Vision*, vol.73, no.2, pp.139-157, 2007.

[6] D. A. Ross, J. Lim, R. S. Lin, and M. H. Yang, “Incremental learning for robust visual tracking”, *International Journal of Computer Vision*, vol.77, no.1-3, pp.125-141, 2008.

[7] Y. G. Jin, and F. Mokhtarian, “Variational particle filter for multi-object tracking,” *in Proc. IEEE 11th International Conference on Computer Vision (ICCV)*, Rio de Janeiro, Brazil, pp.1-8, 2007.

[8] T. Pinkiewicz, R. Williams, J. Purser, “Application of the particle filter to tracking of fish in aquaculture research,” *in Proc. Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, pp. 457-464, 2008.

[9] T. Gao, Z. H. Liu, J. Zhang, “Redundant Discrete Wavelet Transforms based Moving Object Recognition and Tracking”, *Journal of Systems Engineering and Electronics*, vol. 20, no. 5, pp. 1115-1123, 2009.

[10] N. Otsu, “A Threshold Selection Method from Gray-Level Histogram”, *IEEE Trans.SMC.*, vol. 9, no. 1, pp. 62 – 66, 1979.

[11] T. Gao, Z. H. Liu, S. H. Yue, J. Q. Mei, and J. Zhang, “Traffic Video based Moving Vehicle Detection and Tracking in the Complex Environment”, *Cybernetics and Systems: An International Journal*, vol. 40, no. 7, pp. 569–588, 2009.

[12] L. G. David, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[13] C. Belcher, and Y. Du, “Region-based SIFT approach to iris recognition,” *Optics and Lasers in Engineering*, vol.47, no.1, pp.139-147, 2009.

[14] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A Tutorial on Particle Filters for On-Line Nonlinear/ Nongaussian Bayesian Tracking”, *IEEE Trans. Signal Process.*, vol. 50, no. 2 pp. 174-188, 2002.

[15] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, "Robust tracking-by-detection using a detector confidence particle filter," *In: Proceedings of IEEE International Conference on Computer Vision*, Kyoto, Japan, 1515-1522, 2009.

Tao Gao received his PhD degree in Detection Technology and Automatic Equipment from Tianjin University in 2010. He is interested in artificial intelligence, IntelliSense, internet of physical objects, automatic identification, real-time localization, digital information processing, motion detection, computer vision, cloud computing, and so on. His previous research explored motion tracking, moving object detection, and video based motion pattern reorganization.

His current position is Research Engineer, Electronic Information Products Supervision and Inspection Institute of Hebei Province, Shijiazhuang, China. And he is the author or co-author of more than 30 refereed international journal and conference papers, covering topics of intelligent multimedia processing, intelligent transportation systems and internet of things. He is a member of IEEE, IEEE Consumer Electronics Society, IEEE Computer Society, Association for Computing Machinery, China Computer Federation, and Chinese Institute of Electronics. He is the guest Reviewer of several refereed international journals, i.e., IEEE Transactions on Intelligent Transportation Systems, Multimedia Tools and Applications, Journal of Supercomputing, EURASIP Journal on Advances in Signal Processing, Journal of the Chinese Institute of Engineers, and International Journal of Image Processing, etc. He is TPC member of refereed conferences, e.g., EMEIT 2011, ICEOE 2011, ICCDA 2011, MINES 2010, and ITSIC 2010, etc.

Ping Wang is currently working as a professor in the School of Electrical Engineering and Automation at Tianjin University, Tianjin, China. She graduated from Tianjin University and obtained her master's degree from Tianjin University. In recent years, she has been enrolled in many projects, such as the "National Natural Science Foundation of China" project, the Tianjin Natural Science Foundation project, and some transverse subjects.

Chengshan Wang is currently working as a professor in the School of Electrical Engineering and Automation at Tianjin University, Tianjin, China. He is an executive director of the Chinese Society of Electrical Engineering, review team member of the National Natural Science Foundation of Electrical discipline, and the Yangtze River Scholar.

Zhenjing Yao received her PhD degree in Detection Technology and Automatic Equipment from Tianjin University in 2010. Now she is a docent in department of disaster prevention instrument, Institute of Disaster Prevention Science and Technology. She is interested in ultrasonic signal processing, artificial intelligence, automatic identification, real-time localization, and so on.