# Object Tracking Based on Camshift with Multi-feature Fusion

Zhiyu Zhou
College of Information, Zhejiang Sci-Tech University, Hangzhou, China
Email: zhouzhiyu1993@163.com

Dichong Wu
Business Administration College, Zhejiang University of Finance & Economics, Hangzhou, China
Email: 1601441815@qq.com

Xiaolong Peng
College of Information, Zhejiang Sci-Tech University, Hangzhou, China
Email: wheelo@163.com

Zefei Zhu
College of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou, China
Email: zzf.3691@163.com

Kaikai Luo
College of Information, Zhejiang Sci-Tech University, Hangzhou, China
Email: anhuiluokai@foxmail.com

***Abstract*—It is very hard for traditional Camshift to survive of drastic interferences and occlusions of similar objects. This paper puts forward an innovative tracking method using Camshift with multi-feature fusion. Firstly, SIFT features and edge features of the Camshift in RGB space are counted to reduce the probability of disruption by occlusion and clutter. Then, the texture features are collected to resolve the problems of analogue interference, the texture similarity between current frame and previous frames are calculated to determine the object area. The paper also describes the GM(1,1) prediction model, which could solve the occlusion problems in a novel way. Finally, through the motion trajectory, it can anticipate the exact position of the object. The results of several tracking tasks prove that our method has solved problems of occlusions, interferences and shadows. And it performs well in both tracking robustness and computational efficiency.**

***Index Terms*—Camshift, GM(1,1), SIFT features, Object tracking**

## I. Introduction

Mean Shift [1-2] has been widely used in the field of object tracking, yet lacking of the necessary model updates, in addition of the fixed window width of kernel function, both of them will have evil impacts upon the accuracy of tracking. Camshift [3] appears as an improved algorithm which could automatically adjust the window size to fit the object, thus solving the troubles of scale variations. However, the traditional Camshift exists the following disadvantages:

- The tracking results could be very easily disrupted by similar objects.
- The lack of motion behavior prediction, which might lead to a failure of object tracking, especially when the objects move too fast, running beyond the searching scale of the previous frames, or being interfered by other objects.
- The searching window would shrink to a tiny point when the tracking objects get lost.
- Tracking results are susceptible to the illumination changes.

Dozens of scholars have shown interests on the solutions of these problems. In order to overcome the impacts of scale variations and partial occlusions, Shen Xuanjing [4] introduced a kernel density estimation through a model of color distribution, tracking the moving objects with the image moments by Camshift. Li Chao [5] overcame occlusions and information losses of image likewise, but more elegantly, he proposed a face tracking method based on Haar features detection, improving Camshift algorithm with a weighted histogram probability model. In order to conquer the influence of analogue interferences and scene illumination changes, Sun Hongguang [6] combined the color and motion information together into the Camshift, tracking objects adaptively with an optimized particle filter. Wang Zhaowen [7] put forward a Camshift guided particle filter for object tracking, and he incorporated the particles into the probabilistic framework for proposal distribution. In

Manuscript received April 29, 2013.
Corresponding author: Zhiyu Zhou


doi:10.4304/jsw.9.1.147-153

order to handle the issues of fast movement and strong background disturbances about objects, Wang Xin[8] embedded the improved Camshift algorithm into particle filter, he redistributed the random particle samples, whereby the objects could move toward their maximal posterior probability density. However, particle filter needs collecting considerable amount of particles to approximate complex filtering distribution throughout the entire sampling process, which might give rise to a tremendous computation complexity. Li Jianhong [9] got rid of similar color interference of objects with robustness, who put forth an improved algorithm with SURF. And Exner David [10] presented a multi-histogram accumulation to accomodate complex object changes, handling ambiguous cases of partial or full occlusions. Chu Hongxia [11] combined Camshift with frame difference, can track motion object instantly and accurately. Liang Juan [12] has improved the Camshift effect by a large scale with the Kalman filter, which was used to estimate initial parameters of moving object and adjust iteratively to approach location of each objects. Based on an improved Camshift and Kalman filter, Peng Juanchun [13] proposed a real-time hand tracking system for humanoid robot with a stereo vision method.

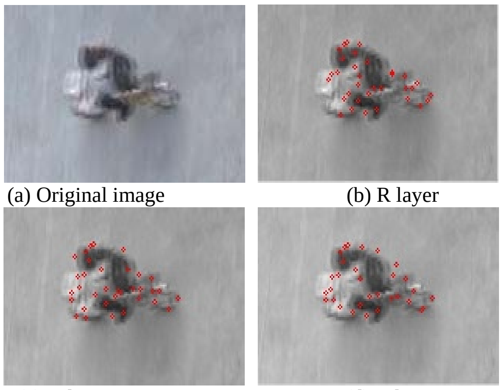
Kalman filter presumes smoothness in relevant motion, which makes possible the process of the modeling in a minor state space, as well as the search in a minor region, maintaining a reliable robustness under the liner time-invariant conditions. Yet in practice, nonlinear situations occur now and then, the deficiency of Kalman filter would expose little by little. However, in the grey system, the internal structures, features and parameters are supposed as unknown. Through a grey generation by limited external behavior of clutter data, the randomness of data would be alleviated. In this paper, we will use grey differential equation to reflect the difference information and harvest the GM(1,1) model [14]. In the actual environment of object tracking, we might meet with the situation of impoverishment and uncertainty in the data information, so as to propagate robust results, it comes necessary to merge grey system theory into the Camshift.

In this paper, edges, textures and SIFT features are fused into the CamShift algorithm. Firstly, we apply Canny operators for the edges detection and extract parameters from the co-occurrence matrix as the texture features in the quantifying process. And then we extract three-layer SIFT features in RGB and track the objects with the fusion of CamShift and GM(1,1) model prediction. The experiment has a satisfactory result that, we can solve the disturbance and the occlusion of similar objects in a robust way.

## II. SIFT Features Extraction

SIFT appears as a superior algorithm standing loftily at the pinnacle of computer vision [15]. Traditional SIFT draws out interest points in the scale space of binary image first. At each candidate points, it will determine location and scale of the fancy key points through stability. And then orientations will coin into each key point as well, based on local image gradient directions. In the end, descriptors will be selected around the whole key points, which allows for robust levels of environment changes.

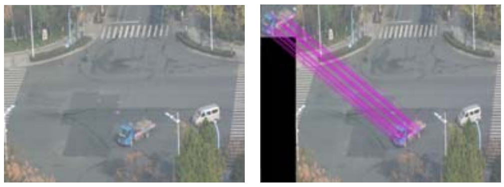
Different from the traditional SIFT, in this paper we will extract the features separately on three channels of RGB space, in such a way that we will get more plentiful features of avail, thus increasing reliability of probability distribution. Figure 1 describes the SIFT features in three-layer space of RGB.



(a) Original image (b) R layer

(c) G layer (d) B layer

Figure1 SIFT features in RGB three-layer space



(a) Original image (b) Matching of target and candidate target

Figure 2 Matching of SIFT features between target and the candidate target

When the features in RGB space generate, we can use Euclidean distance of descriptors as the image similarity criterion. We extract one key point in the first image, finding two closest key points in another image. By comparing the distance of the closest neighbor to that of the second-closest neighbor, we can get a ratio which will thus denote the exact matching points. Figure 2 shows the matches between the tracking object and the candidate object, for the rigid objects in the vehicle, they are saturated with SIFT features.

## III. GM(1,1) Model

The grey system is a new discipline founded by Prof. Deng Julong, it stands out recently owing to its excellent predicting performance, even under the condition of barren information. The GM(1,1) model is defined as followings[14]:

Suppose a non-negative sequence:

$X(0)=\{x^{(0)}(1), x^{(0)}(2),\ldots,x^{(0)}(n)\}$,

Let $x^{(1)}(t)=\sum_{i=1}^{t} x^{(0)}(i),\ t=1,2,\cdots,n$. Such that we get accumulative sequence:

$X^{(1)}=\{x^{(1)}(1), x^{(1)}(2),\ldots, x^{(1)}(n)\}$. Then let

$$z^{(1)}(k)=\frac{1}{2}[x^{(1)}(k)+x^{(1)}(k-1)],\ \mathrm{k}=2,3,\ldots,\mathrm{n} \quad (1)$$

where $Z^{(1)}$ is the $X^{(1)}$ adjacent mean generating sequence, $Z^{(1)}=(z^{(1)}(1),z^{(1)}(2),\cdots,z^{(1)}(n))$, we name:

$$x^{(0)}(k)+az^{(1)}(k)=b \quad (2)$$

as the grey differential equation, also known as GM(1,1) model. A simple form of the differential equations is:

$$\frac{dx^{(1)}}{dt}+ax^{(1)}(k)=b \quad (3)$$

where a and b are the undetermined parameters in the modeling process. Suppose $\hat{a}=[a,b]^T$ are the parameters, and

$$Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix},\quad B=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (4)$$

then parameter estimation of the least square in the GM(1,1) model: $x^{(0)}(k)+az^{(1)}(k)=b$ will satisfy:

$$\hat{a}=(B^TB)^{-1}B^TY \quad (5)$$

Once we get the parameter $\hat{a}$, we are capable to predict the data, the prediction formula is:

$$x^{(0)}(k)=(\beta-\alpha x^{(0)}(1))e^{-a(k-2)} \quad (6)$$

where $\beta=\frac{\mathrm{b}}{1+0.5\mathrm{a}},\alpha=\frac{a}{1+0.5a}$.

We utilize the overall centroids of previous four frames in objects, anticipating their possible position in the following fifth frame. The parameters in GM(1,1) model iterate gradually, such that we guarantee the accuracy of the GM(1,1) model. For the sake of boosting real-time performance of the system, we might as well reduce the computational complexity. To take the x, y coordinates into account respectively, we track objects in either X or Y direction using concurrent GM(1,1) model, descending tracking dimension into one, in such a way that we ameliorate computational complexity of the system.

## IV. CAMSHIFT ALGORITHM

For the traditional Mean Shift, the lack of model updates, as well as the fixed window width of kernel function often has negative effects on the tracking accuracy. As an upgraded version of Mean Shift, Camshift could adjust the window automatically to fit the size of the object, particularly competent for tracking objects with great variations in scale. The major steps of traditional Camshift tracking are as follows：First, it converts the tracking image into HSV space. Through the back projection of color histogram, it can get the probability distributions of colors. At last, by calculating the zero-order moment and first-order moment of image, it will obtain centroid position and the object window size.

The zero-order moment of Camshift is defined as:

$$M_{00}=\sum_x\sum_y I(x,y) \quad (7)$$

The two first-order moments are:

$$M_{10}=\sum_x\sum_y xI(x,y) \quad (8)$$

$$M_{01}=\sum_x\sum_y yI(x,y) \quad (9)$$

The centroid position of searching window defined as:

$$x_c=\frac{M_{10}}{M_{00}},\quad y_c=\frac{M_{01}}{M_{00}} \quad (10)$$

And the size of search window is:

$$s=2\times\sqrt{\frac{M_{00}}{256}} \quad (11)$$

Aiming at the disadvantages of traditional Camshift, in this paper, we will propose an improved Camshift with multiple features fusion. The innovations of our research are as follows:

(1) Instead of utilizing the single H channel in HSV space as the color probability, we will extract the objects information in the channels of RGB.

(2) We will use the SIFT features as the complement of color distribution. As most of features fall upon the objects, we can obtain color probability better via their statistic pixel values.

(3) In either case of existing barren features or existing vast mismatching features, the accuracy of image probability statistics would decrease drastically. Thereby we combine the edge and SIFT features together and extract the color probability distribution from the fusing results to solve these troubles.

(4) Through a fusion of the Camshift and the texture features, we overcome the interference problems of similar objects.

(5) Rather than adopting the true values, we will employ a GM(1,1) model prediction to enhance the anti-occlusion ability when the objects get occluded.

## V. OBJECTS TRACKING BASED ON GM(1,1) AND CAMSHIFT WITH MULT-FEATURE FUSION

Following are the major stages of our computation used to track objects:

(1)To preprocess the image first, and then we are to acquire the edges of objects with the Canny operators.

(2)To extract the SIFT features in three channels of RGB space. Through the statistical feature descriptors we will acquire the accurate size and location of the current objects.

(3) We analyze the pixels of all features, and through the probability histogram of features we could update the objects and recover the original image by the back projection technique.

(4) We utilize the Camshift to track the objects in image sequences continuously. And then we could figure out the exact positions of objects through a comparison of the

texture features between current frames and the previous, judging whether there exist similar objects in color around the objects or not.

(5)If the objects shrink or lose focus suddenly, yet within the range of a frame, we'll predicate them as being occluded. With the attachment of GM(1,1) model, we could predict their exact position in the next frame. At last, we are going to seek the objects in a searching window, whose position lies at center point of prediction and whose size is about 1.5 times of objects in current frame.

## VI. Experimental Results and Analysis

### A. Tracking of maneuvering Object

In this paper, we first get the difference image by the interframe difference and extract the contours by the

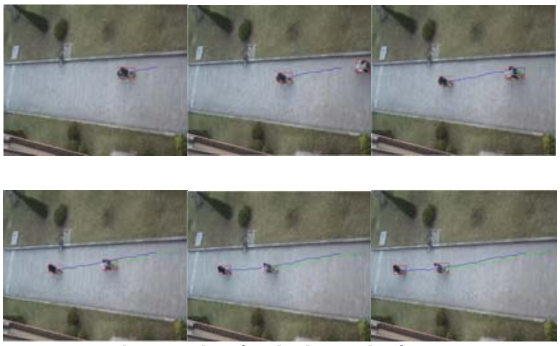


Figure 3 Tracking results of multiple people (frame 20, 35, 50, 65, 80, 95)

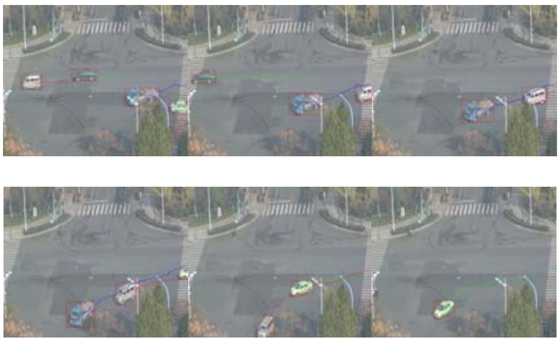


Figure 4 Tracking results of multiple vehicles (frame 50, 65, 80, 95, 125, 140)

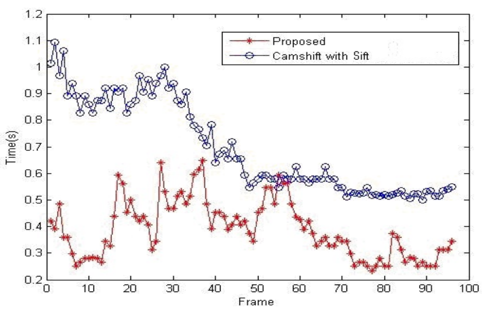



Figure 5 Contrast diagram of traditional Camshift and our algorithm

thresholding binary image, then we draw out the SIFT features from the R, G and B space respectively within the region of contours. The statistical pixel values of SIFT features around the real objects are exploited to calculate the color histogram, and we recover the color probability distribution of original image with the back projection. And then we would acquire center point position and size of the objects in the color probability distribution by the Camshift. After that, we utilize GM(1,1) model to predict the position of the objects in the next frame with the trajectory history, the results manifest that the searching time has been shortened, whereby the real-time performance of the system has increased accordingly. The tracking results of the multiple people are shown in figure 3, and the results of vehicles at crossroads are as shown in figure 4. Despite of a cluster of dynamic changes in vehicles, i.e. the turning or the brake condition, etc., we can still get a stable tracking result with the Camshift. The real-time contrast diagram of the traditional Camshift and our algorithm (figure 4) is shown in figure 5. The experiment results have registered an enhancement in real-time performance of the system.

### B. Object tracking with Occlusions

When we track the objects, a frequent situation we will encounter is the occlusions. In this case, the object area will shrink suddenly or even disappear. When we get the original image by back projection, the areas ratio of probability histogram of current frame to the previous frames will decrease. Similar to this situation, when the Camshift's applied, the center position and the size of the object probability distribution will both grow smaller, once the objects get occluded completely, the Camshift would become invalid at all, which doomed to bring out a tracking failure. Figure 6 shows the tracking results of traditional Camshift algorithm with the SIFT.

As it can be seen from figure 6, without the occlusions, we could track the objects precisely by the traditional Camshift, but as soon as the object's occluded, the true object would easily get tangled with what is not. In this paper, we will assess the occlusions based on measures of whether the objects shrink and disappear. And then we introduce a RGB color probability distribution, owing to the variations of RGB color, the distributions of differed objects might be discrepant as well, thus the tracking objects are non-interfering. When processing the frame, the size and shape of image will be measured, we determine the occlusions by judging whether the ratio of the mean value of them in current frames and previous frames is less than one threshold, in specific, when the ratio of two consecutive frames and the previous frames is consistently less than a threshold, we will discriminate objects as the occlusions. In this sense, we will take advantage of the single GM(1,1) model prediction other than Camshift, and will never modify the size of the object until we track one afterwards. Figure 7 shows the results of our algorithm and marks the trajectory of moving objects in addition.

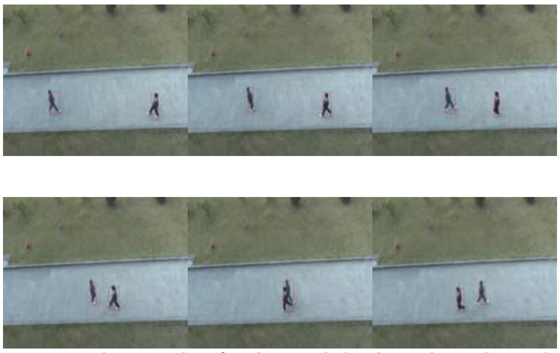


Figure 6 Tracking results of multi-people by the traditional Camshift with occlusions (frame 72, 103, 134, 165, 196, 227)

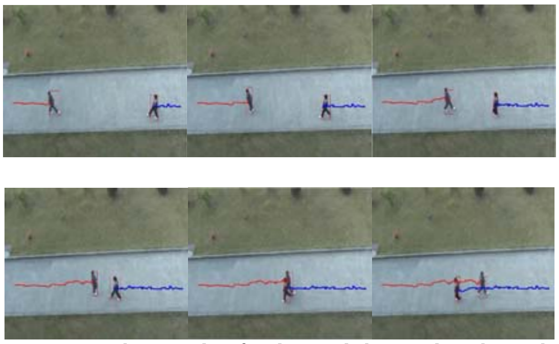


Figure 7 Tracking results of multi-people by our algorithm with occlusions (frame 72, 103, 134, 165, 196, 227)

Figure 8 Tracking results of multi-vehicles by the traditional Camshift with occlusions (frame 65, 76, 87, 98, 109, 120)

Figure 9 Tracking results of multi-vehicles by our algorithm with occlusions (frame 65, 76, 87, 98, 109, 120)

In the process of monitoring the vehicles, similarly, the occlusions occur now and then, i.e. there is a guideboard or other obstacles standing in front. Figure 8 shows the tracking results of traditional Camshift when existing occlusions. And figure 9 shows the tracking results of our algorithm with the occlusions. Figure 10 is the error contrast diagram of vehicle sequences between traditional Camshift with SIFT and our algorithm, the results have proved our algorithm with a higher accuracy.

*C. Objects tracking with the Interference of Similar Objects*

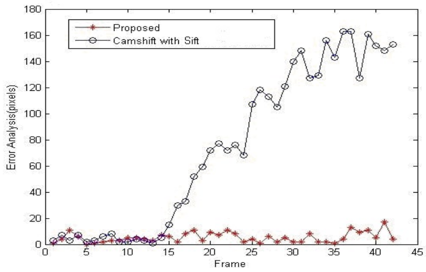



Figure 10 Error contrast diagram of figure 8 and figure 9

When monitoring the track of multiple vehicles in the intersection, some similar vehicles often emerge as the outside interference to spoil the whole color space. While two vehicles of similar colors meet each other halfway, the traditional Camshift algorithm will identify different cars as the one, figure 11 shows the tracking results of traditional Camshift with the interferences. Once some distractions interfere around the tracking objects, the precision of color probability distribution will decrease in response. Thus the features of textures are merged into the algorithm and we can distinguish different objects by textures similarity. After that, the GM(1,1) model is also complemented to solve the occlusion problems. Figure 12 shows the tracking results of vehicles with the occlusions.

*D. Object tracking with Shadows*

In a natural environment (i.e., the shaded outdoor), the context of shade would be severely troublesome for object tracking. As a practical example, we take portions of human body in shade using the traditional Camshift. As is shown in figure 13, the sunshine reflection would blur tracking window, causing numerous mismatches in the process of tracking. A contrast experiment with our algorithm is shown in figure 14. It turns out that the improved algorithm bears an extraordinary robustness, which could survive of strong shadow interferences.

The benefits of improved Camshift algorithm are summarized as follows:

1) Instead of employing color probability of single H component in HSV space, a three-layer RGB space would be utilized otherwise. As the abundance of RGB space, available features in specific class of objects will be saturated.

2) As background information appears to be simple and invariable. SIFT algorithm will also imposed in the calculation of color probability. When SIFT draws out features in image sequences, the majority of features will fall on the real target, color probability would be acquired by the statistic features pixel.

3) The accuracy of probability histogram would be decreased when the features of target are barren and the mismatches occur now and then. In this sense, edge features and the features in the scale of SIFT will be fused as a whole. Through the color histogram of the joint

features, the mismatches have decreased by a large margin.

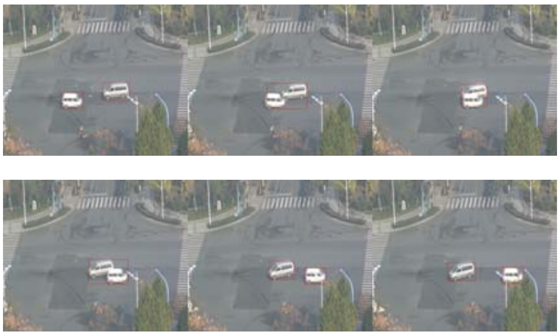


Figure 11 Tracking results of multi-vehicles by the traditional Camshift with interferences (frame 15, 20, 25, 30, 35, 40)

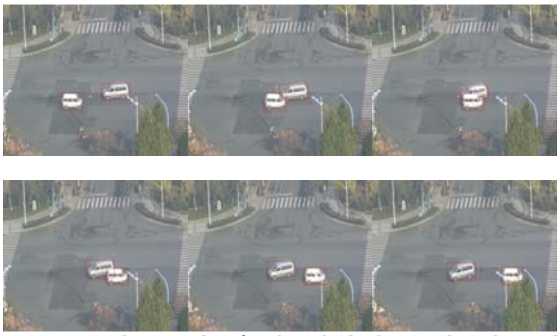


Figure 12 Tracking results of multi-vehicles by our algorithm with interferences (frame 15, 20, 25, 30, 35, 40)

Figure 13 Tracking results of human by the traditional Camshift with shadows (frame 10, 20, 30, 40, 50, 60)

Figure 14 Tracking results of human by our algorithm with shadows (frame 10, 20, 30, 40, 50, 60)

4) Once the interferences of color analogue occur around the target, texture features are attached to recognize the target. Through a combination of Camshift and textures, the analogue interferences have been alleviated to certain degree.

5) When moving objects get occluded, the size of image stays invariable. GM(1,1) model will be kept utilizing to predict the motion position of next frame until the target emerges. Searching in the position of filter prediction, searching time has been saved largely.

## VII. Conclusion

In this paper, we first improve the traditional Camshift, instead of extracting the features in single H channel, we extract the features in all RGB channels with the SIFT, the statistical pixel values are calculated to get the probability distribution, and together with the edges features of object, we improve the precision of the probability distribution by a large margin. When the tracking process is interfered by similar objects, the texture features of the historical frames and the current frames will be compared to determine the exact tracking position. And then on a basis of the continuity of moving objects, we utilize the GM(1,1) model to anticipate the objects' best position in the next frame, narrowing the searching scale sharply. The GM(1,1) model will keep the continuity of the tracking, behaving rather serviceable when existing the occlusions. The experiment results demonstrate our algorithm with a splendid performance of robustness and instantaneity.

## Acknowledgment

This work is supported by Zhejiang Provincial Natural Science Foundation of China (No.LY13F030013), and Teacher's development projects of visiting scholar with higher education in Zhejiang Education Department (No.FX2012012).

## References

[1] Comaniciu Dorin, Ramesh Visvanathan, and Meer Peter, "Real-time tracking of non-rigid objects using Mean Shift," *Proc of the IEEE Conf on Computer Vision and Pattern Recognition*, pp.142-149, 2000.

[2] Comaniciu Dorin, Ramesh Visvanathan, and Meer Peter, "Kernel-based object tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v25, n5, pp.564-577, 2003.

[3] Bradski Gray R, "Computer Vision Face Tracking For Use in a Perceptual User Interface," *Intel Technology Journal*, v 2, n2, pp.1-15,1998.

[4] Shen Xuanjing, and Zhang Bo, "CamShift tracker based on image moments," *Journal of Beijing University of Technology*, v 38, n 1, p 105-109, 2012.

[5] Li Chao, Liu Tiegen, Liu Hongli, et al., "Face tracking based on Haar detection and improved Camshift algorithm," *Journal of Optoelectronics Laser*, v 22, n 12, p 1852-1856, 2011.

[6] Sun Hongguang,Zhang Jin,Liu Yantao,et al., "Optimized Particle Filter Tracking by CamShift Based on Multi-feature," *Opto-Electronic Engineering*, v37, n2, pp 1-6,31,2010.

[7] Wang Zhaowen, Yang Xiaokang, Xu Yi, et al., "CamShift guided particle filter for visual tracking," *Pattern Recognition Letters*, v 30, n 4, pp.407-413, 2009.

[8] Wang Xin, and Tang Zhenmin, "An improved camshift-based particle filter algorithm for real-time target tracking," *Journal of Image and Graphics*, v15, n10, pp.1507-1515, 2010.
[9] Li Jianhong, Zhang Ji, Zhou Zhenhuan, et al., "Object tracking using improved Camshift with SURF method," *2011 IEEE International Workshop on Open-Source Software for Scientific Computation*, pp.136-141, 2011.
[10] Exner David, Bruns Erich, Kurz Danielet, et al., "Fast and robust CAMShift tracking," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.9-16, 2010.
[11] Chu Hongxia, Ye Shujiang, Guo Qingchang, et al., "Object tracking algorithm based on camshift algorithm combinating with difference in frame," *Proceedings of the IEEE International Conference on Automation and Logistics*, ICAL 2007, pp.51-55, 2007.
[12] Liang Juan, Hou Jianhua, Xiang Jun, et al., "An effective automatic tracking algorithm based on Camshift and Kalman filter," *Proceedings of SPIE*, v8003, 2011.
[13] Peng Juanchun, Gu Lizhong, and Su Jianbo, "Hand tracking for humanoid robot using Camshift algorithm and Kalman filter," *Journal of Shanghai Jiaotong University*, v 40, n 7, pp.1161-1165, 2006
[14] Liu Sifeng, Dang Yaoguo, and Fang Zhigeng, "Grey theory and applications," *Third Edition. Beijing: Science Press*, 2004.
[15] Lowe David G, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, v60, n2, pp.91-110, 2004

**Zhiyu Zhou** received B.S. and M.S. degree from Zhejiang Sci-Tech University in 1997 and 2004, respectively. He is now an associated professor at Zhejiang Sci-Tech University and mainly engaged in computer vision, grey systems and robotic tracking. He has published dozens of research papers, 13 papers among them have been included in EI with the first author. He enjoys a profession of guest reviewer in Chinese "Journal of image and graphics" and has been awarded the outstanding reviewer in 2010. He has just accomplished a Natural Science Fund Project of Zhejiang Province - "Target Vision Analysis based on Grey Model" (No.Y1090256), and now presided over another provincial Natural Science Fund Project - "Research on Key Technology of Robotic Autonomous Navigation based on Stereo Vision in Natural Environment" (No.LY13F030013).