Research article

Feature-fusion based object tracking for robot platforms

Xuguang Zhang
Key Lab of Industrial Computer Control Engineering of Hebei Province, Institute of Electrical Engineering, Yanshan University, Qinhuangdao, People's Republic of China

Honghai Liu
Intelligent Systems and Robotics Group, School of Creative Technologies, University of Portsmouth, Portsmouth, UK, and

Yanjie Wang
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, People's Republic of China

Abstract
Purpose – Object tracking has been a challenging problem of robot vision over the decades, which plays a key role in a wide spectrum of visual tracking-related applications such as surveillance, visual servoing, sensing and navigation in robotics, video compression. The purpose of this paper is to present a novel intensity, orientation codes and geometry (IOCG) histogram variant of the mean-shift algorithm for object tracking.
Design/methodology/approach – Feature cues of intensity, orientation codes and geometric information are fused together to form an IOCG histogram in combination with a conventional mean-shift-based tracking algorithm.
Findings – Experimental results demonstrate the effectiveness and efficiency of the proposed method. Not only do fusing orientation codes features allow the proposed algorithm to conduct tracking in a cluttered background, but partial occlusion is also solved in the tracker in that spatial information usually lost in a conventional histogram is compensated by the introduced geometric relations between tracked pixels and the center of a tracker template.
Originality/value – The paper presents a novel vision tracking method for robots.

Keywords Robotics, Tracking, Object-oriented methods, Control technology, Statistical testing

Paper type Research paper

1. Introduction

Object tracking is a key component in robot vision system such as surveillance (Kettnaker and Zabih, 1999), visual servoing (Marchand and Chaumette, 2005), video compression (Bue *et al.*, 2002), driver assistance (Handmann *et al.*, 1998) and perceptual user interfaces (Bradski, 1998), etc. It has been recognized as one of the bottlenecks in visual tracking-related research and industrial applications in that most applications require robustness characteristic of object tracking algorithms subject to partial occlusion, clutter, environmental lighting changes, etc. To cope with these difficulties, contributions had been made to precise object tracking in the past two decades. Tracking algorithms can be broadly classified into target representation and localization and filtering and data association (Comaniciu *et al.*, 2003; Sinha *et al.*, 2007). The former are bottom-up approaches which cope with changes of appearance of a target; the latter are top-down approaches which handle dynamics of objects of interest, learning of scene priors and evaluation of different hypotheses. The target representation is composed of feature-based tracking (Beymer *et al.*, 1997; Wang *et al.*, 2009), contour-based tracking (Isard and Blake, 1996; Nascimento and Marques, 2004; Usabiaga *et al.*, 2009) and region-based tracking (Hager and Belhumeur, 1998). In region-based trackers, either a histogram or non-parametric descriptions are employed to represent the object of interest (Adam *et al.*, 2006; Jeong and Jaynes, 2008); it is evident that the mean-shift approach (Comaniciu *et al.*, 2003; Collins, 2003; Juan and Hu, 2008; Tu *et al.*, 2009) has been given priority in recent research progress. However, the mean-shift algorithm has some shortcomings for feasible tracking applications. First, the conventional mean-shift algorithm lacks acceptable performance in gray-scale images or infrared (IR) images in that a singular gray space representing the targets

The current issue and full text archive of this journal is available at **www.emeraldinsight.com/0143-991X.htm**

Industrial Robot: An International Journal
38/1 (2011) 66–75

[DOI 10.1108/01439911111097869]

This research was supported by National Natural Science Foundation of China (Grant Nos. 60805045, 60804036) and UK Engineering and Physical Sciences Research Council (Grant No. EP/G041377/1). The authors also thank Professor. F. Porikli for providing the database of IR image sequences.

usually leads to false positions for the scenarios whose background is in a similar gray style. Second, loss of spatial information is inherent in the use of histograms. Traditional histograms are popular representations for non-parametric density, but they disregard the spatial arrangement of related feature values. Finally, localizing objects of interest might be inaccurate for scenarios in which there exist occlusions.

We propose a novel intensity, orientation codes and geometry (IOCG) histogram method to represent the features of an object of interest with an aim to improve the capability of the conventional mean-shift tracker in different tracking scenarios in gray-scale images. First, we integrated the orientation codes into a histogram to improve the distinguishing capability of a template in a cluttered background. Then, we employed the geometric relations between the pixels and the center of its template to overcome partial object occlusion in that it takes into account the spatial information usually lost in traditional histograms. Next, we initialized the conventional mean-shift tracker at multiple positions to avoid falling into the local extrema during tracking. Finally, we employed the log-likelihood ratio between an object of interest and its background to efficiently estimate the position of the object, which significantly reduces the computational cost. The reminder of the paper is organized as follows: Section 2 proposes an IOCG histogram consisting of individual features cues, which is used to represent an object of interest. Section 3 reviews the mean-shift algorithm; Section 4 presents multi-position initialization for the tracker. Section 5 validates the proposed algorithms in examples of object tracking on real video sequences. Finally, Section 6 concludes the paper.

2. IOCG histogram

An IOCG histogram is proposed in this section, the histogram has the capability of representing an object of interest in gray images in terms of individual features cues including its intensity, orientation codes and geometrical positions. Such a histogram can be presented mathematically in equation (1) as follows:

$$\hat{p}(u,v,s) = \sum_{(i,j)\in I} \delta(u - c_{ij})\delta(v - I_{ij})\delta(s - G_{ij}) \qquad (1)$$

where δ is the Kronecker delta function. $u \in (0, \ \ldots, \ n-1)$ is the index of the orientation codes feature cues, $v \in (0, \ \ldots, \ m-1)$ is the index of intensity feature cues and $s \in (0, \ \ldots, \ g-1)$ is the index of the geometrical position. c_{ij}, I_{ij}, G_{ij} are the values of three features for one pixel in an image region. Hence, the proposed histogram can be employed to represent an object of interest in that the histogram can be constructed in terms of IOCG features which are the combined feature description of the object of interest.

2.1 Intensity feature

Intensity feature is crucial in object tracking in that it is easy to describe the robustness subject to dynamic shape of a target of interest. However, it is evident that the higher the accuracy of the intensity feature, the more unstable the performance of the intensity histogram subject to fluctuations in illumination. Hence, intensity feature in 16 bins is employed for gray-scale images in order to reduce the computational cost and to fluctuations in illumination. For example, when the illumination changes from 255 to 252, the intensity histogram with 16 bins sets all of these to 15.

2.2 Orientation codes

For discrete images, the orientation codes (Ullah and Kaneko, 2004) are obtained as quantized values of gradient angles around each pixel by applying a differential operator into horizontal and vertical derivatives like Sobel operator, further calculating their ratio can be employed to obtain its gradient angle. The orientation codes are shown in Figure 1; given an interval of the gradient angle $\Delta\theta$, the orientation codes for a pixel location (i, j) c_{ij} can be obtained as follows:

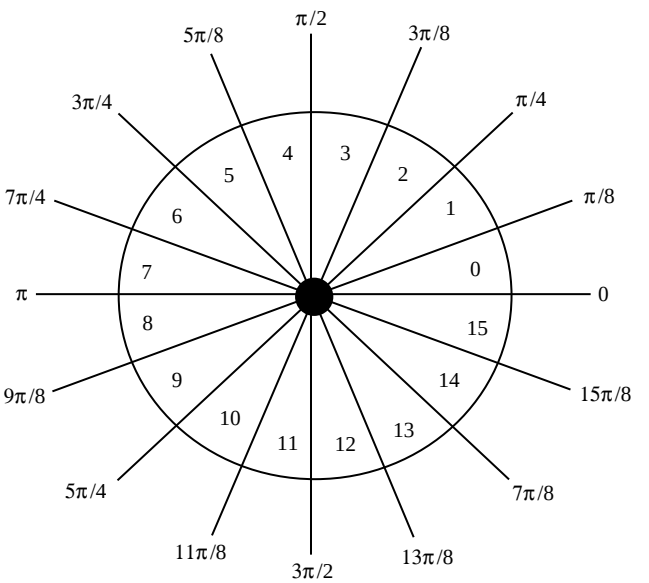
$$c_{ij} = \begin{cases} \left[\frac{\theta_{ij}}{\Delta\theta}\right] & \left|\frac{\partial f}{\partial x}\right| + \left|\frac{\partial f}{\partial y}\right| > \Gamma \\ m & otherwise \end{cases} \qquad (2)$$

where $\theta_{ij} = \tan^{-1}[(\partial f/\partial y)/(\partial f/\partial x)]$, $\partial f/\partial y$ and $\partial f/\partial x$ are the gradient of vertical and horizontal orientation computed by the Sobel operator. Though the feature of the orientation codes represents edge information of an image region, the orientation codes extracted from low-contrast regions could not provide rich edge information. In order to obtain stable extraction of the orientation codes features, the pixels are discarded if they do not meet the threshold value Γ, i.e. $|\partial f/\partial x| + |\partial f/\partial y| < \Gamma$. Additionally, a trade-off has to be decided when choosing the number of orientation codes due to the fact that the larger the number of the orientation codes, the richer the information they carry and the more sensitive rotation changing is. Further threshold value Γ is selected based on the richness of the gradient of an object of interest. Based on our empirical results and the lessons learned from Ullah and Kaneko's (2004) study, the orientation codes are defined as an integer range from 0 to 15 (i.e. $\Delta\theta$ is set as $\pi/8$ radians). The threshold value Γ is set as 5 by default in this paper.

2.3 Geometric relation

Since the intensity and orientation codes are only used to represent statistic properties of an object of interest, spatial

Figure 1 Illustration of the orientation codes



information, such as the distribution of the pixels of the object, has to be taken into account. In our work, spatial information is represented by the geometric relation between the pixels and the center of the target model, which has been taken into account in the histogram, based on the intensity and orientation code features. An object of interest can be represented by a histogram fusing IOCG features. The geometry restriction is shown in Figure 2, in order to describe the geometry feature of pixel P(x, y); this pixel is related to the center of the corresponding image region. Further, the length of this line is used to represent the distance parameter of point P, and the angle between this line and x-axis is employed to represent the angle parameters. On the other hand, in order to achieve the robustness and insensitivity of the parameters subject to the disturbance of slight rotation, the geometry features are quantified as 16 bins, representing 16 regions divided by four quadrants and four ring regions in an image region. Finally, spatial information $G(i,j)$ of pixel (i,j) is integrated into the histogram of $G(i,j) = A(i,j) \times 4 + D(i,j)$, where A represents the quantification of angle parameter and D represents the quantification of distance parameter.

2.4 IOCG histogram-based occlusion solving

It is quite common that an object of interest is often occluded by another object during tracking. The proposed method is compared with intensity histogram and spatial intensity histogram proposed by Xu *et al.* (2005) in order to evaluate the performance of the IOCG histogram in terms of coping with partial occlusion of a target.

An example is given in Figure 3, where the target was initialized in Figure 3(a), and Figure 3(b) shows a frame where the target was occluded by a book. Figure 3(c) is the template image extracted from Figure 3(a). Five regions were randomly selected from Figure 3(b), i.e. Figure 3(d)-(h), then we match the template with five other regions using

Figure 2 Geometric relation between the pixels and the centre of an object of interest

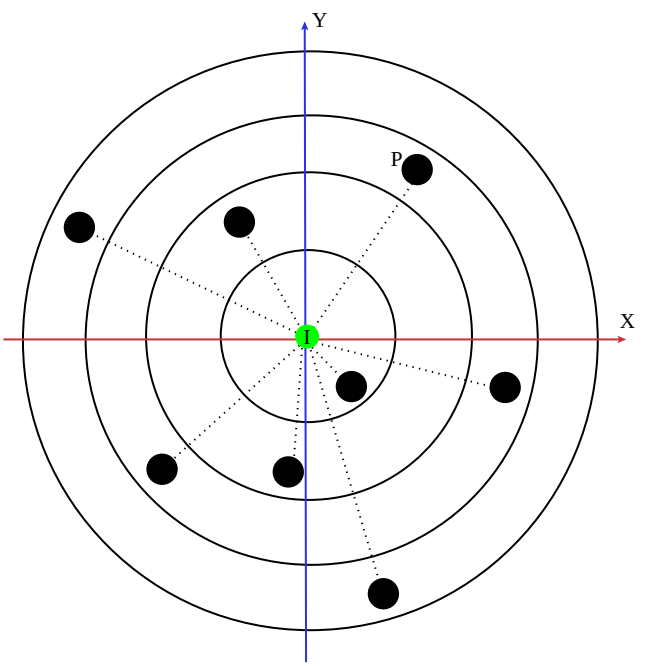
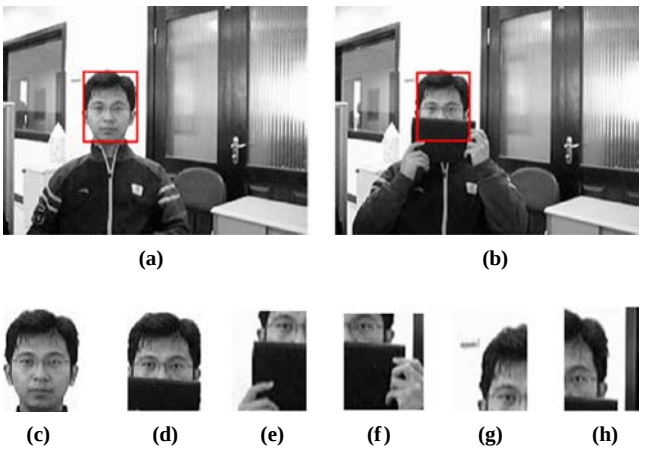


Figure 3 Accurate or illusive regions of an occluded object of interest



Notes: Region (c); the target template; region (d); the tracking region and regions (e)-(h); the randomly selected false regions

Bhattacharyya coefficient. The results, provided in Table I, demonstrates that the IOCG histogram gains the maximal similarity at the accurate region (d); intensity and spatial intensity histograms gain the maximal similarity at the illusive regions (g) or (h). Hence, it is evident that the IOCG histogram can deal with the tracking tasks better than the other two histogram models.

3. Mean-shift algorithm

The mean-shift algorithm (Fukanaga and Hostetler, 1975) is a non-parametric technique that climbs the gradient of a probability distribution to search the nearest dominant mode. It has been increasingly employed in real-time object tracking field due to its computational efficiency and non-parametric mode-seeking characteristic (Comaniciu *et al.*, 2000).

For a set of sampling data, $\{x_i\}_{i=1,\ldots,n} \in R^d$, the multivariate kernel density estimate with kernel $K(x)$ is given by:

$$\hat{f}(x) = \frac{1}{h^d} \sum_{i=1}^{n} w_i \cdot K\left(\frac{x - x_i}{h}\right) \tag{3}$$

where x denotes the d-dimensional variable, w_i denotes the ith weight and h the window radius, also known as bandwidth parameter.

The core of the mean-shift algorithm is a gradient descent optimization method in which the mean-shift vector M_h is iteratively computed to move the location vector x to a new location $x_1 = x + M_h(x)$ along it. The mean-shift vector M_h is derived by:

Table I An object occlusion scenario using different feature models

Position	IOCG histogram	Intensity	Spatial intensity
(d)	0.59	0.91	0.75
(e)	0.44	0.89	0.77
(f)	0.53	0.90	0.81
(g)	0.58	0.94	0.82
(h)	0.48	0.94	0.85

$$M_h(x) = \frac{\sum_{i=1}^{n} x_i w_i g(\|(x - x_i)/h\|^2)}{\sum_{i=1}^{n} w_i g(\|(x - x_i)/h\|^2)} - x \quad (4)$$

where $g(x) = -k'(x)$ and $k(\|x\|^2) = K(x)$ is the profile function of kernel K(x) (Ning Song *et al.*, 2005).

Convergence properties of the mean-shift method are crucial in feasible object tracking. Cheng (1995) claimed that mean shift is an instance of gradient ascent with an adaptive step size It is discovered that, unlike Newton's method, each iteration of the mean shift is guaranteed to bring us closer to a stationary point; on the other hand, it also shares the weakness of Newton's method, the mean shift may get a wrong local maximum during the process of iteration (Fashing and Tomasi, 2005). One of the methods which is used to avoid these discontinuities is to take small steps for moving the direction of the local gradient. Unfortunately, if the step size is too large, the rate of convergence cannot be guaranteed (Comaniciu and Meer, 2002). Concerning the trade-off between practical object tracking, multi-position initialization is proposed to avoid local maxima during algorithm convergence.

4. Object tracking

Object tracking is the process of locating objects of interest in time in image sequences. The outputs are the location of the object of interest within the image sequences; usually, it is assumed that information of the objects of interest is known at the initial frame. An object tracking algorithm usually consists of prior information capture and data filtering. In the proposed work, the IOCG histogram in Section 3 is employed as a feature template to represent an object of interest O; the conventional mean shift is used to incorporating prior information represented by the IOCG into image frame filtering. During tracking, we can assume that we have a previous estimate of the position from previous frame, and we will initial the mean-shift tracker at this estimate in current image. Hence, the contribution of the tracking algorithm is to locate the position which minimizes the similarity function in formula (7).

4.1 Mean-shift tracker

The mean-shift tracker can be divided into three steps. First, the PDFs of object model $\hat{q}_{uvs}$ and object candidate $\hat{p}_{uvs}(y)$ are calculated based on the IOCG histogram and kernel density estimation as follows:

$$\hat{q}_{uvs} = C_q \sum_{i=1}^{n_q} k\left(\left|\frac{x_i}{h_q}\right|^2\right) \delta[b_u(x_i) - u]\,\delta[b_v(x_i) - v]\,\delta[b_s(x_i) - s], \quad (5)$$
$$u = 1, 2, \ldots, m - 1,\ v = 1, 2, \ldots, n - 1,$$
$$s = 1, 2, \ldots, g - 1$$

$$\hat{p}_{uvs}(y) = C_p \sum_{i=1}^{n_p} k\left(\left|\frac{y - x_i}{h_p}\right|^2\right) \delta[b_u(x_i) - u]\,\delta[b_v(x_i) - v] \times \delta[b_s(x_i) - s], \quad (6)$$
$$u = 1, 2, \ldots, m - 1,\ v = 1, 2, \ldots, n - 1,$$
$$s = 1, 2, \ldots, g - 1$$

where k is the kernel function, h_q and h_p are the size of both kernel, b_u, b_v and b_s are index functions for features of IOCG bins, respectively. In this paper, m, n, g are set as 16, respectively. C_q and C_p are the normalization factors. δ is the Kronecker delta function. In order to increase the robustness of the density estimation, smaller weights are assigned to the pixels which are farther away from the center since the peripheral pixels have less reliability, which are often substantially affected by occlusions or interference of the background. The Epanechnikov kernel is adopted in this case.

Second, the similarity measure Bhattacharyya coefficient is employed to locate the position which is the closest to the IOCG-based template. The Bhattacharyya coefficient (Comaniciu *et al.*, 2000) is defined as:

$$\hat{\rho}(y) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} \sum_{s=0}^{g-1} \sqrt{\hat{p}_{uvs}(y)\hat{q}_{uvs}} \quad (7)$$

Finally, the object tracking problem can be reduced as the problem of maximizing $\hat{\rho}(y)$ around a neighborhood of the object position y_0 at previous frame, which can be solved by an iterative mean-shift procedure as follows:

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_p} x_i w_i g(\|(\hat{y}_0 - x_i)/h_p\|^2)}{\sum_{i=1}^{n_p} w_i g(\|(\hat{y}_0 - x_i)/h_p\|^2)} \quad (8)$$

The iteration is stopped if the mean-shift iterative is converged or a threshold of the predefined number of iterations is satisfied.

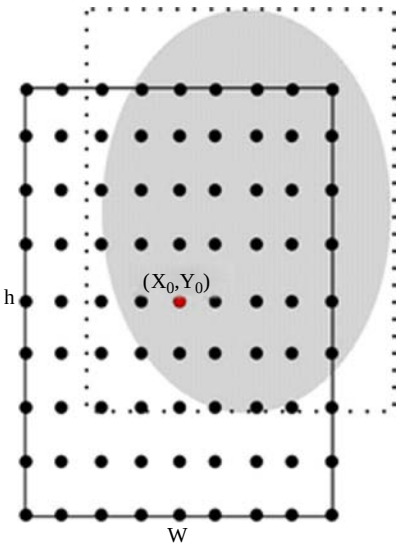
4.2 Adaptive scale selection

Kernel scale is a crucial parameter to the performance of the proposed mean-shift algorithm. If the kernel size is chosen too large, the tracking window might contain redundant background pixels. On the other hand, the best localization might not be obtained if the kernel size is set too small. However, the scale of an object of interest often changes in different image frames in order to achieve better tracking performance, so the kernel size has to be adapted accordingly. The approach in Comaniciu *et al.* (2003) is employed to adapt the kernel size in the tracking algorithm. In this paper, two regions are generated first by enlarging and shrinking the scale of a candidate region by 10 per cent in current frame over the size of a target of interest in its previous frame, then these three regions are employed to match with the target model; finally, the optimal scale is chosen by which gives the highest score in equation (7). Some limitations of this method had been indicated in Collins (2003) and Adam *et al.* (2006). For example, if an object being tracked is uniform in color, then there is a tendency for the kernel size to shrink. In another case of partial occlusion of a target, we confronted with an additional dilemma: if we suppose that a gray target is partially occluded, a good score may be provided by shrinking the target and locating it around the non-occluded part. In order to solve the problems, the orientation codes and geometry information are combined into a histogram in the proposed algorithm. It is demonstrated that the proposed algorithm has better performance on dealing with the partial occlusion and locating the target at the correct position.

4.3 Multiple position initialization

One of the major issues for object tracking is that the trackers are sensitive to noise, lack robustness and can be trapped into local maxima. Multiple position initialization is proposed with an aim to improve their robustness in this paper. The fact is that, while the shading area of an object of interest as shown in Figure 4 is updated, the more involvement of a background and geometric

Figure 4 Initialization of the mean-shift based tracker at multiple positions



relative information, the more likely the tracking window drifts away from the target center or even loses the target. However, if the position of the target at the previous frame is multiple positions (x_0, y_0), the mean-shift tracker is initialized at 81 positions, which are generated in the region from the left-bottom position $(x_0 - w/2, y_0 - h/2)$ to the right-top position $(x_0 + w/2, y_0 + h/2)$ with the steps of w/8 and h/8. (where w and h are the width and height of the template in Figure 4, respectively). Then, a final target position can be obtained at the highest Bhattacharyya coefficient when applying mean-shift iterations at individual positions.

To reduce computation complexity, a log-likelihood method is proposed to evaluate whether a position is valid. It is intended to initialize the positions of the object of interest only. It is shown in Figure 5, where the region within the red rectangle is used to obtain the object's PDF, while the region between the blue and the red rectangles is to obtain the background's PDF (Collins *et al.*, 2005; Babu *et al.*, 2007). Consider the gray feature of an image, the histograms $H_{obj}(i)$ and $H_{bg}(i)$ are applied to describe the probabilities of the ith pixel belonging to the object and background, respectively, where i represents the ith gray level. The log-likelihood of individual pixels within the object of interest and background region is provided as follows:

$$L(i) = \log \frac{\max\{H_{obj}(i), \varepsilon\}}{\max\{H_{bg}(i), \varepsilon\}} \tag{9}$$

where ε is a small non-zero value (i.e. $\varepsilon = 0.001$) which prevents from being divided by zero or taken by the log of zero. The nonlinear log likelihood ratio in equation (9) maps object/background distributions into positive values for intensity distinctive to the object while negative values are marked for the background. To gain practically reliable object pixels, a threshold is set to segment the likelihood image as below:

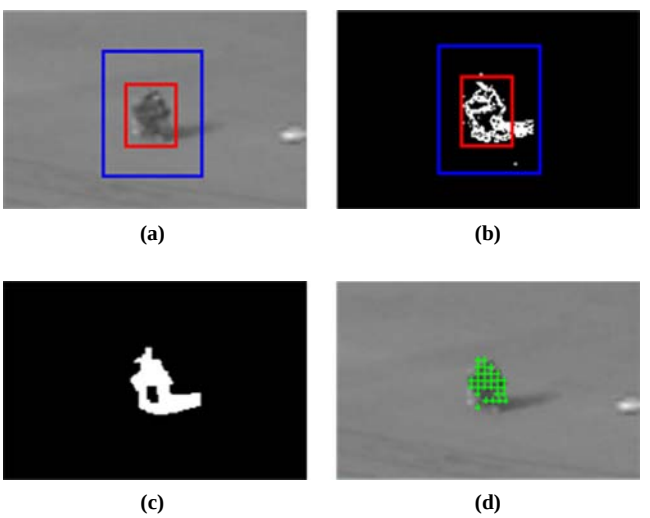
$$I_{\log}(i) = \begin{cases} 1 & \text{if } L(i) > T' \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where T′ is the predefined threshold used for selecting reliable object pixels. Figure 5(b) shows the likelihood image of the object of interest and its background. Since there are many hole areas in the object, morphological operations are employed to eliminate the holes as shown in Figure 5(c), the selected initialized positions are shown in Figure 5(d).

5. Experimental results

The proposed IOCG histogram is evaluated for both rigid and non-rigid targets in image sequences with a focus on gray-scale images. The IOCG histogram is quantized into $16 \times 16 \times 16$ bins, coefficient α is defined to adjust the size relation between an object of interest and its template (i.e. $size_T = \alpha \times size_{obj}$). It is demonstrated that the parameter

Figure 5 Illustration of likelihood map



Notes: (a) Frame with labeled object (red box) and background (blue box); (b) likelihood map of the object and background; (c) processed by the threshold and morphological operations; (d) points initialized (green points)

improves the robustness of the tracker by setting a template a little smaller than the target in that it prevents it from including more background pixels. α is set as 0.8 by default, for instance, an object of interest as shown in Figure 6 is located by a red circle and the template is represented by a blue circle. The first frame of the box sequence as shown in Figure 7 shows that the proposed method is robust to scale change. The learned lesson is that the farther away from a camera and the more complex the background, the less effective the proposed algorithm is. The size of the target in the initial frame is 96×120 pixels and is shrunk to 16×20 pixels in the 450 frame. Figure 8 shows that the proposed method can effectively track an object of interest in a complex background; it is also demonstrated in an occlusion scenario. Note that the algorithm distinguishes the object of interest from a similar object as shown in Figure 8(c). Further,

Figure 6 Initialization of the object of interest (the red ellipse) and the template (the blue ellipse)

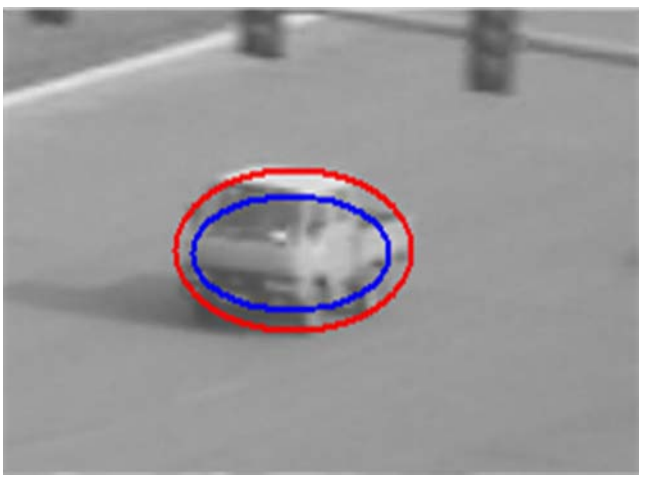
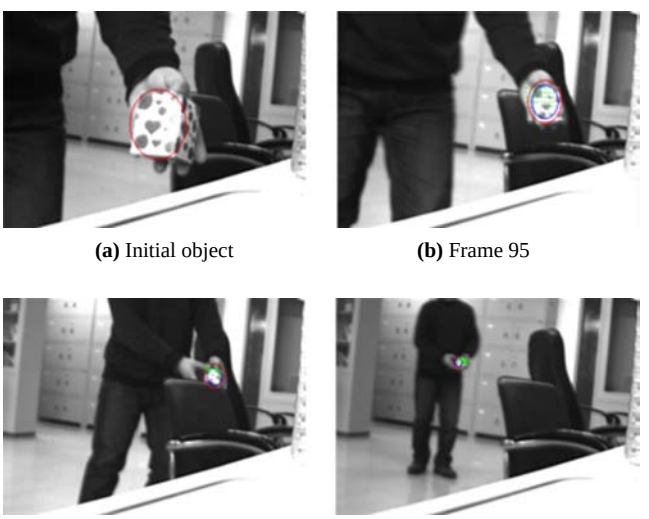


Figure 7 Box sequence: the scale of the target changes

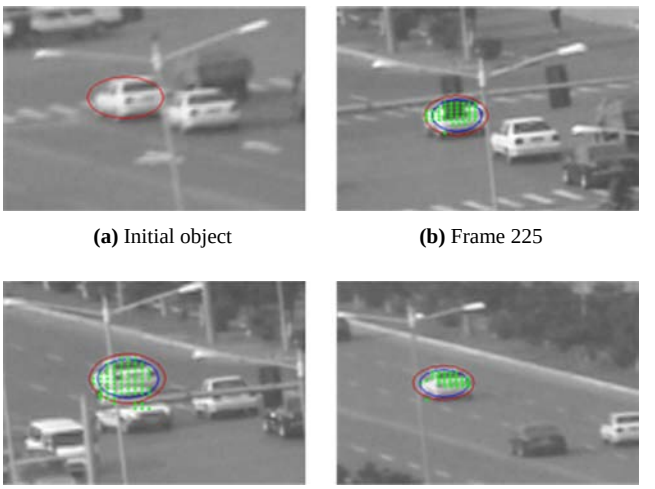


(a) Initial object **(b)** Frame 95

(c) Frame 267 **(d)** Frame 450

Notes: The " + " (green) indicates the initial positions of mean-shift tracker; the red ellipse (object) and blue ellipse (template) show the tracking result

Figure 8 White car sequence: the scale of the target changes. The object was occluded by other object. There are similar car near by the object



(a) Initial object **(b)** Frame 225

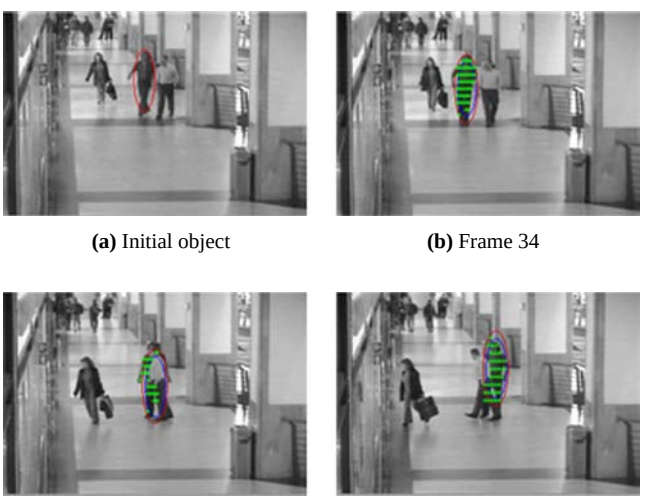
(c) Frame 308 **(d)** Frame 613

Notes: The " + " (green) indicates the initial positions of mean-shift tracker; the red ellipse (object) and blue ellipse (template) show the tracking result

OneShopOneWait2cor sequences from CAVIAR database as shown in Figure 9 are used to present the robustness of the proposed algorithms for image sequences with occlusion and non-rigid objects. The degree of object occlusion is shown in Figure 9(c) and (d); it is convincing that the proposed method has good capability of handling object tracking in serious occlusion.

The proposed algorithm is compared with the conventional mean-shift algorithm and its variant (Xu *et al.*, 2005) in various datasets in order to evaluate its effectiveness. The intensity information is quantized into 16 bins for the conventional algorithm; the intensity information is quantized

Figure 9 OneShopOneWait2cor sequence: the object is nonrigid and is severe occluded by another man



(a) Initial object **(b)** Frame 34

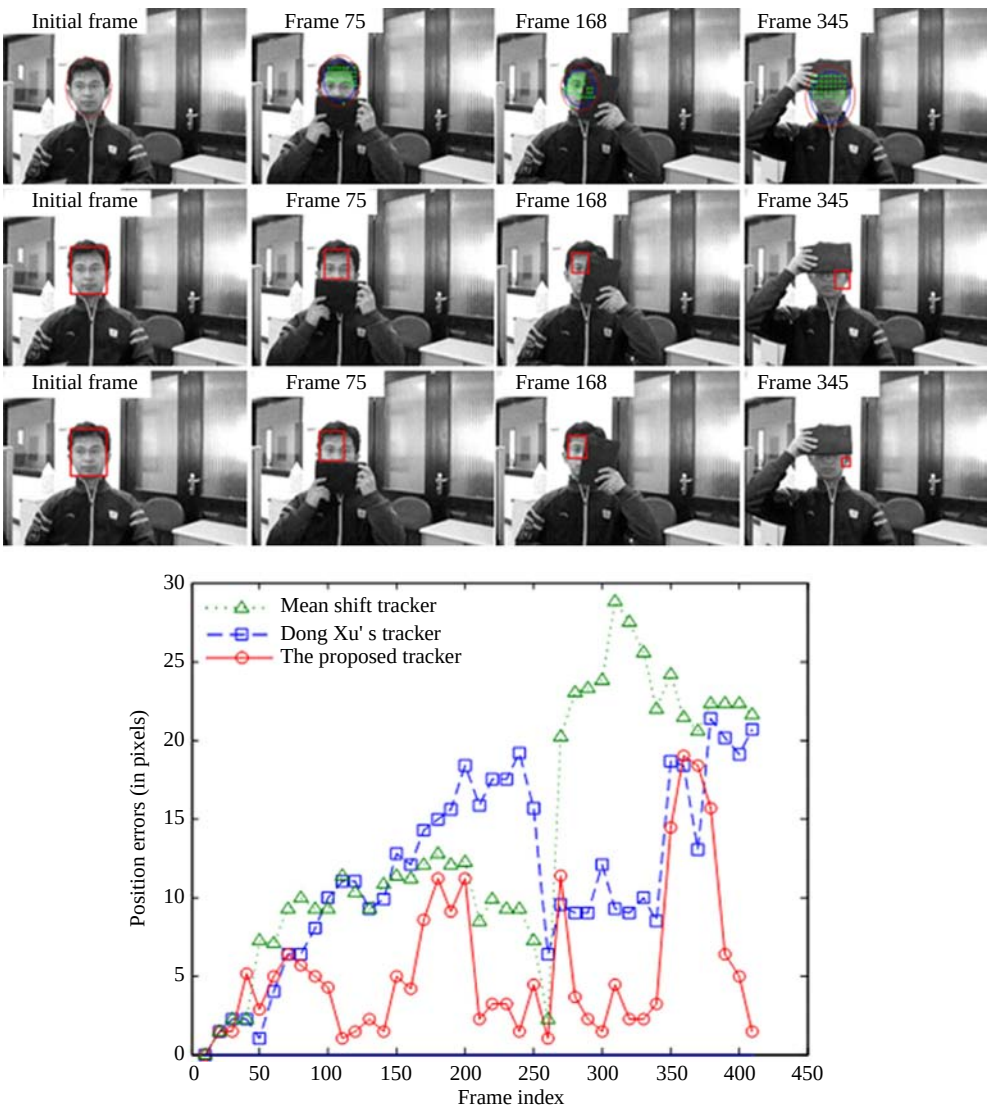
(c) Frame 164 **(d)** Frame 175

Notes: The " + " (green) indicates the initial positions of mean-shift tracker; thfe red ellipse (object) and blue ellipse (template) show the tracking result

into 16 bins and its spatial information into four bins in the experiments. We set the scale of target as the method expatiated in Section 4.2. Figure 10 shows the simulation result of face tracking, where the size of target is 48×64 pixels. Different degrees of object occlusion are provided in image frames 72, 188 and 366; the results clearly demonstrate that the proposed algorithm significantly outperforms the standard mean-shift method and Xu's variant. Further, we compare the results of both algorithms with the ground truth, which are marked manually in an interval of ten frames. The results shown in Figure 10 present that the conventional mean-shift tracker is not as accurate as the proposed method, e.g. it even loses the target in the range around image frame 400. It confirms the outstanding tracking ability of the proposed method. Figure 11 shows the simulation result obtained from *EnterExitCrossingPaths1cor* sequences in CAVIAR database, in which the target is occluded by a subject. The size of target is 26×70 pixels; we compare the results of both algorithms with the ground truth, which are marked manually in an interval of ten frames. It can be seen that, except for a few frames, the proposed method outperforms the conventional mean-shift method and the variant in terms of the position errors.

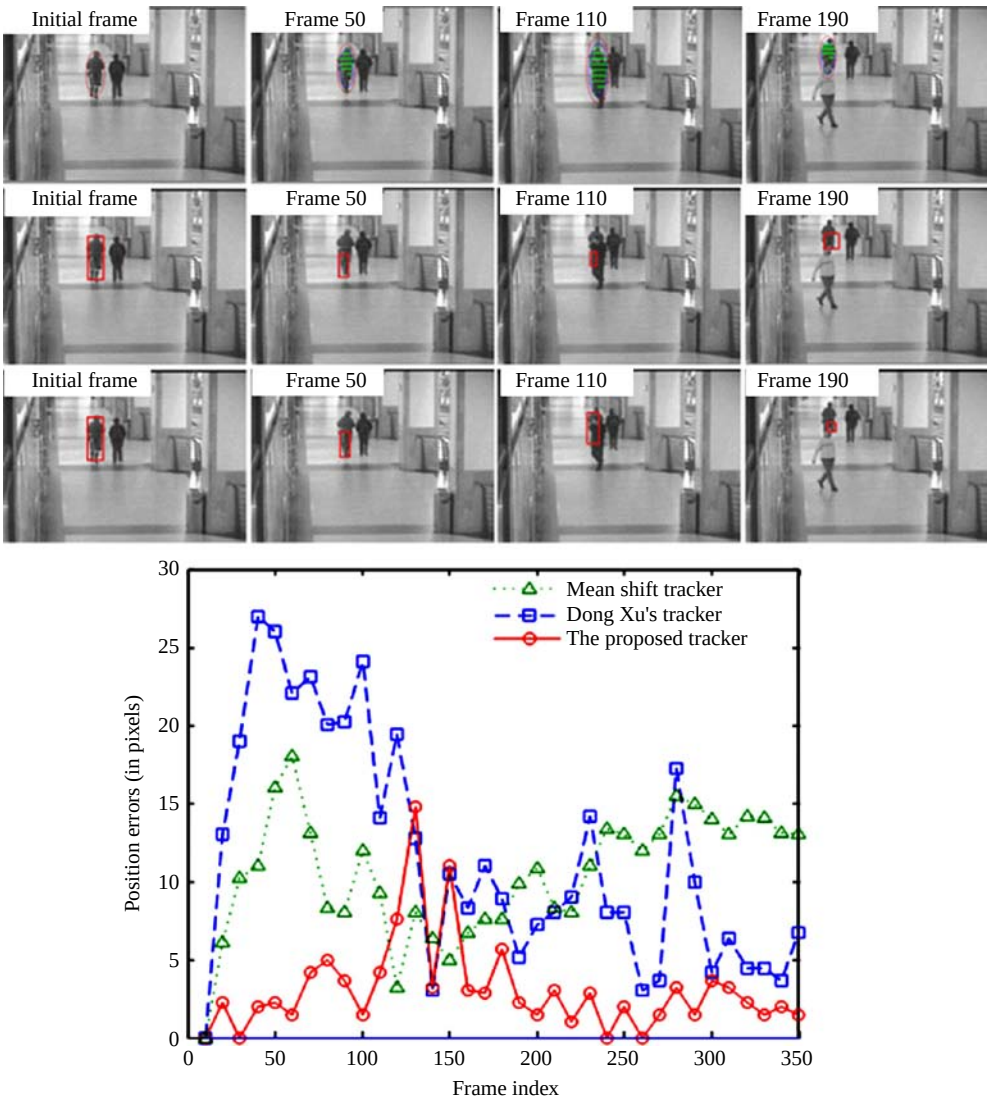
The proposed mean-shift variant is further evaluated on IR image sequence as shown in Figure 12, the target size is 16×40 pixels. The results of both mean-shift methods are compared against the ground truth, marked manually in an interval of ten frames. The results of position errors clearly demonstrate that the proposed method outperforms the two mean-shift tracking algorithms in terms of tracking accuracy.

Figure 10 Position errors' comparison of the proposed algorithm, Xu's algorithm and conventional mean-shift algorithm for Face sequence



Notes: First row: tracking with the proposed tracker; the " + " (green) indicates the initial positions of mean-shift tracker; the red ellipse (object) and blue ellipse (template) show the tracking result; second row: tracking with Dong Xu's tracker; the red rectangle shows the tracking result; third row: tracking with the mean-shift tracker; the red rectangle shows the tracking result; position errors over frame number in face sequences as shown in the last row; symbol "○" denotes the result from our tracker; symbol "□" denotes the result from Dong Xu's tracker; Symbol "Δ" denotes the result from standard mean-shift tracker

Figure 11 Position errors' comparison of the proposed algorithm, Dong Xu's algorithm and conventional mean-shift algorithm for EnterExitCrossingPaths1cor sequence



Notes: First row: tracking with the proposed tracker; the " + " (green) indicates the initial positions of mean-shift tracker; the red ellipse (object) and blue ellipse (template) show the tracking result; second row: tracking with Dong Xu's tracker; the red rectangle shows the tracking result; third row: tracking with the mean-shift tracker; the red rectangle shows the tracking result; position errors over frame number in face sequences as shown in the last row; symbol "○" denotes the result from our tracker; symbol "□" denotes the result from Dong Xu's tracker; symbol "Δ" denotes the result from standard mean-shift tracker

The proposed algorithm was implemented in terms of VC++ program on a 2.8 GHz PC. We initialized the mean-shift tracker with 81 kernels. The computing time is shown in Table II, it demonstrates that the computing time increases as the target size gets bigger; a real-time tracker can be implemented under the condition that the image size of a target is less than 48×48.

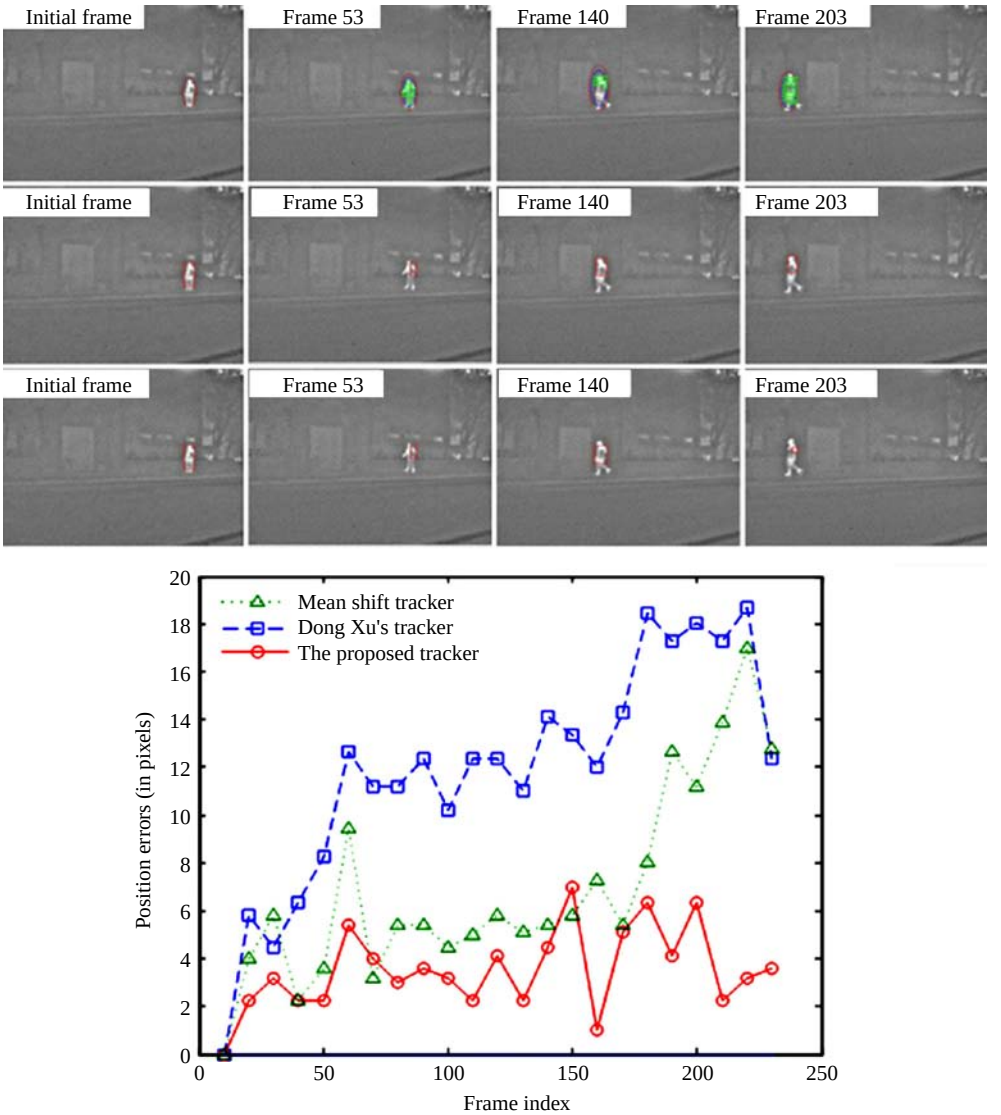
6. Concluding remarks

This paper presented a novel vision tracking method for robots. We have fused feature cues of intensity, orientation codes and geometric information together to form an IOCG histogram in combination with a conventional mean-shift-based tracking algorithm. Experimental results had demonstrated that the effectiveness of the proposed tracking algorithm, whose advantages can be summarized as follows:

- The IOCG histogram embodies both spatial and statistical properties of objects and elegant solution to fuse multiple features.
- The orientation codes and geometry feature make a robust tracking in complex background for gray-scale images.
- Our method is robust to partial occlusions and gains the accuracy localization.
- By using the likelihood map and by initializing the mean-shift tracker at multiple positions, the algorithm can escape local extrema.

Future work is focused on investigating the real-time performance and object tracking robustness of the proposed

Figure 12 Position errors' comparison of the proposed algorithm, Dong Xu's algorithm and conventional mean-shift algorithm for IR sequence



Notes: First row: tracking with the proposed tracker; the " + " (green) indicates the initial positions of mean-shift tracker; the red ellipse (object) and blue ellipse (template) show the tracking result; second row: tracking with Dong Xu's tracker; the red rectangle shows the tracking result; third row: tracking with the mean-shift tracker; the red rectangle shows the tracking result; position errors over frame number in face sequences as shown in the last row; symbol "○" denotes the result from our tracker; symbol "□" denotes the result from Dong Xu's tracker; symbol "Δ" denotes the result from standard mean-shift tracker

Table II Computing time of different target size

Target size (in pixels)	20 × 20	32 × 32	40 × 40	48 × 48	64 × 64	100 × 100
Computing time (ms)	19	23	31	46	66	165

algorithm when it is integrated into applied robotic systems. Background information will be also combined into the tracker to improve its real-time performance.

References

Adam, A., Rivlin, E. and Shimshoni, I. (2006), "Robust fragments-based tracking using the integral histogram", IEEE Conference on Computer Vision and Pattern Recognition, June, pp. 798-805.

Babu, R.V., Pérez, P. and Bouthemy, P. (2007), "Robust tracking with motion estimation and local Kernel-based color modeling", *Image and Vision Computing*, Vol. 25, pp. 1205-16.

Beymer, D., McLauchlan, P., Coifman, B. and Malik, J. (1997), "A real-time computer vision system for measuring traffic parameters", paper presented at IEEE Conference

on Computer Vision and Pattern Recognition, San Juan, Puerto Rico.

Bradski, G.R. (1998), "Computer vision face tracking as a component of a perceptual user interface", *IEEE Workshop Applications of Computer Vision, Princeton, NJ*, pp. 214-19.

Bue, A.D., Comaniciu, D., Ramesh, V. and Regazzoni, C. (2002), "Smart cameras with real-time video object generation", *IEEE International Conference Image Processing, Rochester, NY*, Vol. 3, pp. 429-32.

Cheng, Y. (1995), "Mean shift, mode seeking, and clustering", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 17 No. 8, pp. 790-9.

Comaniciu, D., Ramesh, V. and Meer, P. (2003), "Kernel-based object tracking", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 25 No. 5, pp. 564-75.

Collins, R.T. (2003), "Mean shift blob tracking through scale space", *IEEE International Proceedings Computer Vision Pattern Recognition, Madison, WI*, Vol. 2, pp. 234-40.

Collins, R.T., Liu, Y. and Leordeanu, M. (2005), "Online selection of discriminative tracking features", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27 No. 10, pp. 1631-43.

Comaniciu, D. and Meer, P. (2002), "Mean shift: a robust approach toward feature space analysis", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 24 No. 5, pp. 603-19.

Comaniciu, D., Ramesh, V. and Meer, P. (2000), "Real-time tracking of non-rigid objects using mean shift", *Proceedings of the IEEE International Computer Vision Pattern Recognition, Hilton Head Island*, Vol. 2, pp. SC142-9.

Fashing, M. and Tomasi, C. (2005), "Mean shift is a bound optimization", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27 No. 3, pp. 471-4.

Fukanaga, K. and Hostetler, L.D. (1975), "The estimation of the gradient of a density function, with application in pattern recognition", *IEEE Transactions on Information Theory*, Vol. 21, pp. 32-40.

Kettnaker, V. and Zabih, R. (1999), "Bayesian multi-camera surveillance", IEEE Conference Computer Vision and Pattern Recognition, June 23-25, 1999, Fort Collins, CO, pp. 253-9.

Hager, G. and Belhumeur, P. (1998), "Efficient region tracking with parametric models of geometry and illumination", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 20 No. 10, pp. 1125-39.

Handmann, U., Kalinke, T., Tzomakas, C., Werner, M. and Seelen, W.V. (1998), "Computer vision for driver assistance systems", *SPIE*, Vol. 3364, pp. 136-47.

Isard, M. and Blake, A. (1996), "Contour tracking by stochastic propagation of conditional density", European Conference on Computer Vision, Cambridge, pp. 343-56.

Jeong, K. and Jaynes, C. (2008), "Object matching in disjoint cameras using a color transfer approach", *Journal of Machine Vision and Applications*, Vol. 19 Nos 5-6, pp. 443-55.

Juan, C.W. and Hu, J.S. (2008), "A new spatial-color mean-shift object tracking algorithm with scale and orientation estimation", 2008 IEEE International Conference on Robotics and Automation, Pasedena, CA, pp. 1-6.

Marchand, E. and Chaumette, F. (2005), "Feature tracking for visual servoing purposes", *Robotics and Autonomous Systems*, Vol. 52 No. 1, pp. 53-70.

Nascimento, J.C. and Marques, J.S. (2004), "Robust shape tracking in the presence of cluttered background", *IEEE Translations on Multimedia*, Vol. 6 No. 6, pp. 852-61.

Ning Song, P., Jie, Y. and Zhi, L. (2005), "Mean shift blob tracking with kernel histogram filtering and hypothesis testing", *Pattern Recognition Letters*, Vol. 26, pp. 605-14.

Sinha, S.N., Frahm, J.M., Pollefeys, M. and Genc, Y. (2007), "Feature tracking and matching in video using programmable graphics hardware", *Journal of Machine Vision and Applications* (online).

Tu, J., Tao, H. and Huang, T. (2009), "Online updating appearance generative mixture model for mean shift tracking", *Journal of Machine Vision and Applications*, Vol. 20 No. 3, pp. 163-73.

Ullah, F. and Kaneko, S. (2004), "Using orientation codes for rotation-invariant template matching", *Pattern recognition*, Vol. 37, pp. 201-9.

Usabiaga, J., Erol, A., Bebis, G., Boyle, R. and Twombly, X. (2009), "Global hand pose estimation by multiple ellipse tracking", *Journal of Machine Vision and Applications*, Vol. 21 No. 1, pp. 1-15.

Wang, J., Bebis, G., Nicolescu, M., Nicolescu, M. and Miller, R. (2009), "Improving target detection by coupling it with tracking", *Journal of Machine Vision and Applications*, Vol. 20 No. 4, pp. 1432-769.

Xu, D., Wang, Y. and An, J. (2005), "Applying a new spatial color histogram in mean-shift based tracking algorithm", *Proceedings of the Image and Vision Computing*, University of Otago, Dunedin, pp. 1-6.

Corresponding author

Xuguang Zhang can be contacted at: xuguang.zhang78@gmail.com