International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 23, No. 7 (2009) 1245–1263

ROBUST OBJECT TRACKING USING JOINT COLOR-TEXTURE HISTOGRAM*

JIFENG NING†

College of Information Engineering
Northwest A&F University
Yangling, Shaanxi, P. R. China
jf_ning@sina.com

LEI ZHANG‡ and DAVID ZHANG§

‡*Biometrics Research Center*
Department of Computing
Hong Kong Polytechnic University
Kowloon, Hong Kong, P. R. China
‡*cslzhang@comp.polyu.edu.hk*
§*csdzhang@comp.polyu.edu.hk*

CHENGKE WU

†*State Key Laboratory of Integrated Service Networks*
Xidian University, Xi'an, P. R. China
ckwu@xidian.edu.cn

A novel object tracking algorithm is presented in this paper by using the joint color-texture histogram to represent a target and then applying it to the mean shift framework. Apart from the conventional color histogram features, the texture features of the object are also extracted by using the local binary pattern (LBP) technique to represent the object. The major uniform LBP patterns are exploited to form a mask for joint color-texture feature selection. Compared with the traditional color histogram based algorithms that use the whole target region for tracking, the proposed algorithm extracts effectively the edge and corner features in the target region, which characterize better and represent more robustly the target. The experimental results validate that the proposed method improves greatly the tracking accuracy and efficiency with fewer mean shift iterations than standard mean shift tracking. It can robustly track the target under complex scenes, such as similar target and background appearance, on which the traditional color based schemes may fail to track.

Keywords: Object tracking; mean shift; local binary pattern; color histogram.

*This work was supported by the Hong Kong Polytechnic University Internal Grant (A-SA08), and the National Science Foundation Council of China under Grants 60532060 and 60775020
‡Author for correspondence

1. Introduction

Real-time object tracking is a critical task in computer vision applications. Many tracking algorithms[26] have been proposed to overcome the difficulties arising from noise, occlusion, clutter and changes in the foreground object or in the background environment. Among the various tracking algorithms,[26] mean shift tracking algorithms have recently become popular due to their simplicity and efficiency.[3,7,10,13,25]

The mean shift algorithm was originally proposed by Fukunaga and Hostetler[8] for data clustering. It was later introduced into the image processing community by Cheng.[5] Bradski[3] modified it and developed the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm to track a moving face. Comaniciu and Meer successfully applied mean shift algorithm to image segmentation[6] and object tracking.[7] Mean Shift is an iterative kernel-based deterministic procedure which converges to a local maximum of the measurement function with certain assumptions on the kernel behaviors. Furthermore, mean shift is a low complexity algorithm, which provides a general and reliable solution to object tracking and is independent of the target representation.

Currently, a widely used form of target representation is the color histogram,[7,14] which could be viewed as the discrete probability density function (PDF) of the target region. Color histogram is an estimating mode of point sample distribution and is very robust in representing the object appearance. However, using only color histograms in mean shift tracking has some problems.[25] First, the spatial information of the target is lost. Second, when the target has similar appearance to the background, color histogram will become invalid to distinguish them. For a better target representation, the gradient or edge features have been used in combination with color histogram.[7,10] Several object representations that exploit the spatial information have been developed by partitioning the tracking region into fixed size fragments,[1] meaningful patches[12] or the articulations of human objects.[20] For each subregion, a color or edge feature based target model was presented.

The texture patterns,[9,18,22–24,28] which reflect the spatial structure of the object, are effective features to represent and recognize targets. Since the texture features introduce new information that the color histogram does not convey, using the joint color-texture histogram for target representation is more reliable than using only color histogram in tracking complex scenes. The idea of combining color and edge for target representation has been exploited by researchers.[7,10] However, how to utilize effectively both the color intensity and texture features is still a difficult problem. This is because though many texture analysis methods, such as gray concurrence matrices[9] and Gabor filtering,[23] have been proposed, they have high computational complexity and cannot be directly used together with color histogram.

The local binary pattern (LBP)[16,17] technique is very effective to describe the image texture features. LBP has advantages such as fast computation and rotation invariance, which facilitates the wide usage in the fields of texture analysis, image retrieval, face recognition, image segmentation, etc.[2,15,19,21,27,28] Recently,

LBP was successfully applied to the detection of moving objects via background subtraction.[11] In LBP, each pixel is assigned a texture value, which can be naturally combined with the color value of the pixel to represent targets. In Ref. 13, Nguyen *et al.* employed the image intensity and the LBP feature to construct a two-dimensional histogram representation of the target for tracking thermographic and monochromatic video.

In this paper, we adopt the LBP scheme to represent the target texture feature and then propose a joint color-texture histogram method for a more distinctive and effective target representation. The major uniform LBP patterns are used to identify the key points in the target region and then form a mask for joint color-texture feature selection. The proposed target representation scheme eliminates smooth background and reduces noise in the tracking process. Compared with the traditional RGB color space based target representation, it efficiently exploits the target structural information and hence achieves better tracking performance with fewer mean shift iterations and higher robustness to various interferences of background and noise in complex scenes.

The paper is organized as follows. Section 2 briefly introduces the mean shift algorithm. Section 3 analyzes LBP and presents the joint color-texture histogram scheme in detail. Experimental results are presented and discussed in Sec. 4. Section 5 concludes the paper.

2. Mean Shift Algorithm

2.1. *Target representation*

A target is usually defined by a rectangle or an ellipsoidal region in the image. Most existing target tracking schemes use the color histogram to represent the rectangle or ellipsoidal target. In this paper, we will present a new target representation approach by using the joint color-texture histogram. First let us review the target representation in the mean shift tracking algorithm.[7]

Denote by $\{x_i^*\}_{i=1\cdots n}$ the normalized pixel positions in the target region, which is supposed to be centered at the origin point. The target model $\hat{q}$ corresponding to the target region is computed as

$$\begin{cases} \hat{q} = \{\hat{q}_u\}_{u=1\cdots m} \\ \hat{q}_u = C\sum_{i=1}^{n} k(\|x_i^*\|^2)\delta[b(x_i^*) - u] \end{cases} \tag{1}$$

where $\hat{q}_u$ represent the probabilities of feature u in target model $\hat{q}$, m is the number of feature spaces, δ is the Kronecker delta function, $b(x_i^*)$ associates the pixel x_i^* to the histogram bin, $k(x)$ is an isotropic kernel profile and constant C is a normalization function defined by

$$C = 1 \Big/ \sum_{i=1}^{n} k(\|x_i^*\|^2) \tag{2}$$

Similarly, the target candidate model $\hat{p}(y)$ corresponding to the candidate region is given by

$$\begin{cases} \hat{p}(y) = \{\hat{p}_u(y)\}_{u=1\cdots m} \\ \hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta[b(x_i)-u] \end{cases} \tag{3}$$

$$C_h = 1 \Big/ \sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{4}$$

where $\hat{p}_u(y)$ represents the probability of feature u in the candidate model $\hat{p}(y)$, $\{x_i\}_{i=1\cdots n_h}$ denote the pixel positions in the target candidate region centered at y, h is the bandwidth and constant C_h is a normalization function.

In order to calculate the likelihood of the target model and the candidate model, a metric based on the Bhattacharyya coefficient is defined between the two normalized histograms $\hat{p}(y)$ and $\hat{q}$ as follows:

$$\rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y)\hat{q}_u} \tag{5}$$

The distance between $\hat{p}(y)$ and $\hat{q}$ is then defined as

$$d[\hat{p}(y), \hat{q}] = \sqrt{1-\rho[\hat{p}(y), \hat{q}]} \tag{6}$$

2.2. *Mean shift tracking*

Minimizing the distance (6) is equivalent to maximizing the Bhattacharyya coefficient (5). The iterative optimization process is initialized with the target location y_0 in the previous frame. Using Taylor expansion around $\hat{p}_u(y_0)$, the linear approximation of the Bhattacharyya coefficient (5) is obtained as

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0)\hat{q}_u} + \frac{1}{2}C_h \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{7}$$

where

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u\,(\mathrm{y}_0)}}\,\delta[b(\mathrm{x}_i)-u] \tag{8}$$

Since the first term in (7) is independent of y, to minimize the distance in (6) is to maximize the second term in (7). In the iterative process, the estimated target moves from y to a new position y_1, which is defined as

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\|\frac{y-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{y-x_i}{h}\right\|^2\right)} \tag{9}$$

When we choose kernel g with the Epanechnikov profile,[7] (9) is reduced to

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i}{\sum_{i=1}^{n_h} w_i} \tag{10}$$

By using (10), the mean shift tracking algorithm finds in the new frame the most similar region to the object. For more information about color histogram based target representation and mean shift tracking, please refer to Ref. 7.

3. Target Tracking with Joint Color-Texture Histogram

3.1. *Local binary pattern*

The LBP[16,17] operator labels the pixel in an image by thresholding its neighborhood with the center value and considering the result as a binary number (binary pattern). The general version of the LBP operator is defined as follows:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{11}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c) of a local neighborhood and g_p to the gray values of P equally spaced pixels on a circle with radius R. The function $s(x)$ is defined as follows:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{12}$$

Figure 1 is an example of $\mathrm{LBP}_{8,1}$ ($P = 8$, $R = 1$). By varying P and R, we have the LBP operators under different quantization of the angular space and spatial resolution, and multiresolution analysis can be accomplished by using multiple $\mathrm{LBP}_{P,R}$ operators.

The texture model derived by (11) has only gray-scale invariance. The gray-scale and rotation invariant LBP texture model is obtained by (refer to Ref. 16 for

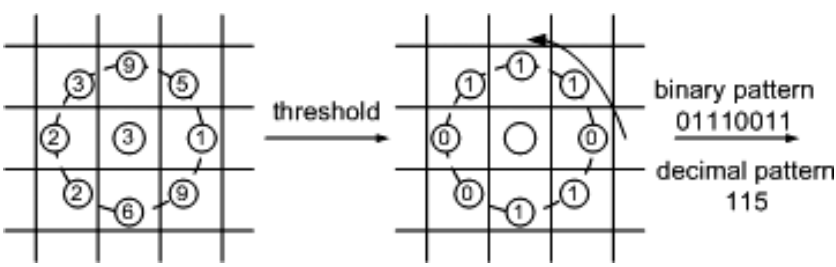


Fig. 1. An example of $\mathrm{LBP}_{8,1}$ texture model.

details)

$$\mathrm{LBP}_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(\mathrm{LBP}_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \tag{13}$$

where

$$U(\mathrm{LBP}_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{14}$$

The superscript "*riu2*" means that the rotation invariant "uniform" patterns have a U value of at most 2. By definition, the $P+1$ "uniform" binary patterns occur in a circularly symmetric neighbor set of P pixels. Equation (13) assigns a unique label to each of them corresponding to the number of "1" bits in the pattern (0 to P), while the "nonuniform" patterns are grouped under the "miscellaneous" label $(P+1)$.

3.2. *Target representation with joint color-texture histogram*

A limitation of LBP is that it does not work robustly on flat regions where the gray values have small fluctuations. In order to make LBP more robust against these subtle changes in pixel values, Heikkiä and Pietikäinen[11] modified the thresholding strategy in the LBP operator by replacing the term $s(g_p - g_c)$ in Eqs. (11), (13) and (14) with $s(g_p - g_c + a)$. The greater the value |a| is, the higher fluctuations in pixel values are allowed without affecting much the thresholding result. In this paper, we adopt this modified thresholding method and employ $\mathrm{LBP}_{8,1}^{riu2}$ to describe the target texture features because of its low computational complexity.

With the above analysis, we calculate the LBP feature of each point in the image region, whose value is between 0 and 9. Thus an appearance model combining the color and texture is constructed and it consists of color channel and LBP texture pattern. However, compared with the usual color based target representation, this direct combination does not enhance much mean shift tracking. Especially, when the target is very similar to background, representing the target by using color and LBP texture features on the whole tracking region is hard to distinguish them due to the strong interference from the background. Therefore, a demand exists to find a new way to combine the color and LBP texture features more effectively.

The $\mathrm{LBP}_{8,1}^{riu2}$ model has nine uniform texture patterns,[16] which are shown in Fig. 2. Each of the $\mathrm{LBP}_{8,1}^{riu2}$ uniform patterns is regarded as a micro-texton [Ref. 4, pp. 198–201]. The local primitives detected by the $\mathrm{LBP}_{8,1}^{riu2}$ model include spots, flat areas, edges, line ends and corners, etc. In Fig. 2, the white circles represent "1" and the black circles represent "0".

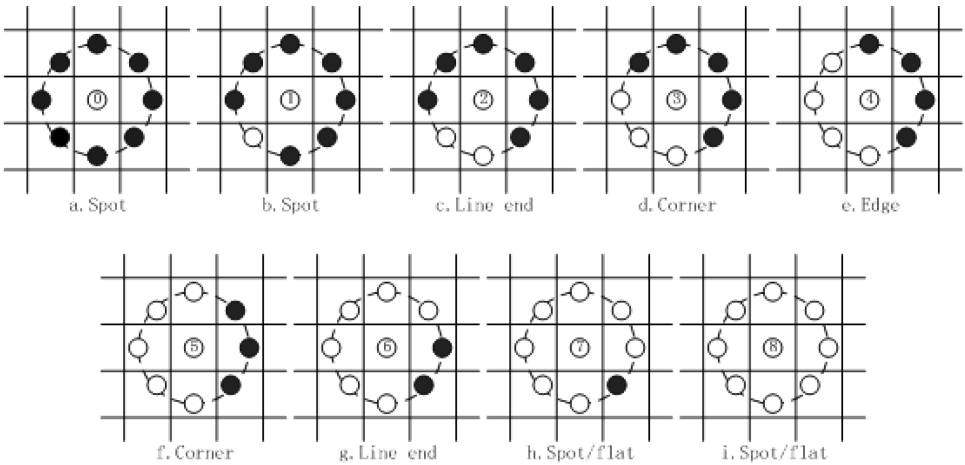


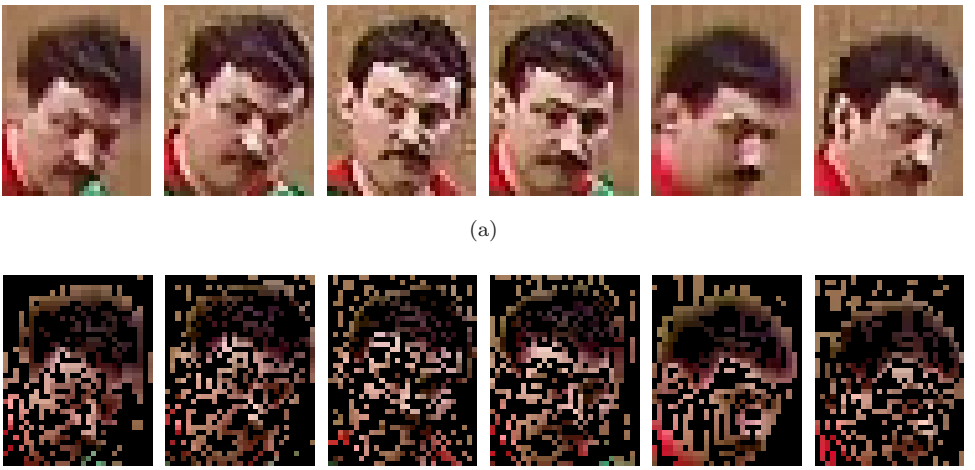
Fig. 2. The nine uniform patterns of $\mathrm{LBP}_{8,1}^{riu2}$ model.

In target representation, the micro-textons such as edges, line ends and corners, by name of "major uniform patterns", represent the main features of target, while spots and flat areas, called "minor uniform patterns", are minor textures. Thus, we extract the main uniform patterns of the target by the following equation

$$\mathrm{LBP}_{8,1}^{riu2} = \begin{cases} \sum_{p=0}^{7} s(g_p - g_c + a) & U(\mathrm{LBP}_{8,1}) \leq 2 \quad \text{and} \\ & \sum_{p=0}^{7} s(g_p - g_c + a) \in \{2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

In $\mathrm{LBP}_{8,1}^{riu2}$, the labels corresponding to minor uniform patterns are 0, 1, 7 and 8 respectively, and the label of nonuniform patterns is 9. The labels corresponding to main uniform patterns are 2–6, which have five patterns. Equation (15) groups the minor uniform patterns as nonuniform patterns. Generally, the main LBP features of a target are more important than its minor features to represent the target. Therefore, by (15) we extract only the pixels corresponding to the main LBP features and then use the color and texture features of these pixels to model the target. That is to say, we first use (15) to form a mask and then use the color and LBP features within this mask to model the target appearance model.

We denote by M1 the original color appearance model in Ref. 7, by M2 the joint color and LBP texture feature model computed by (13) and by M3 the proposed joint color and LBP model using the major pattern mask. Actually, M2 is basically the method proposed by Nguyen *et al.*[23] For the appearance models M1 and M2, all pixels in the target region will be used for object tracking. For the proposed model, only part of the pixels is extracted by (15) to represent the main object features.



(a)

(b)

Fig. 3. (a) The tracking windows and (b) the masks extracted by the LBP major uniform patterns.

With these points, we combine the color histogram with LBP histogram to model the target. The proposed target representation model M3 effectively removes the smoothing background and extracts the main features. It reduces greatly the interferences induced by minor uniform patterns, which come from mainly the smooth background and noise in the target area.

Figure 3 shows an example of the target model mask established by the proposed scheme M3. Figure 3(a) shows a sequence of the target regions in adjacent frames and Fig. 3(b) shows the corresponding masks (nonblack pixels), which include the key feature points in the target region, formed by (15). It is seen that the masks change dynamically with the content changing of the target region. Most of the smooth background pixels and part of the smooth target pixels are eliminated by the masks in object tracking.

3.3. *The tracking algorithm with the joint color-texture histogram*

We use the RGB channels and the LBP patterns extracted by (15) to jointly represent the target and embed it into the mean shift tracking framework. To obtain the color and texture distribution of the target region, we use (1) to calculate the color and texture distribution of the target model $\hat{q}$, in which $u = 8 \times 8 \times 8 \times 5$. The first three dimensions (i.e. $8 \times 8 \times 8$) represent the quantized bins of color channels and the fourth dimension (i.e. 5) is the bin of the modified LBP texture patterns by (15). Similarly, the target candidate model $\hat{p}(y)$ is calculated with (3). The whole tracking algorithm is summarized as follows.

Mean Shift Tracking Algorithm with Joint Color-Texture Histogram

Input: the target model $\hat{q}$ and its location y_0 in the previous frame.

(1) Initialize the iteration number $k \leftarrow 0$.
(2) In the current frame, calculate the distribution of the target candidate model $\hat{p}(y_0)$.
(3) Calculate the weights $\{w_i\}_{i=1\cdots n_h}$ using (8).
(4) Calculate the new location y_1 of the target candidate using (10).
(5) Let $k \leftarrow k+1$, $d \leftarrow \|y_1 - y_0\|$, $y_0 \leftarrow y_1$. Set the threshold ε and the maximum iteration number N.
If $d < \varepsilon$ or $k \geq N$
Stop and go to Step 6.
Otherwise
Go to step 2.
(6) Load the next frame as the current frame with initial location y_0 and go to Step 1.

4. Experimental Results

In this section, extensive and representative experiments are performed to illustrate and testify the proposed joint color-texture model based mean shift tracking algorithm (i.e. M3) in comparison with the mean shift tracking with appearance models M1 and M2. The videos of different scenes, including the one that has similar target/background colors, are used in evaluating the performance of different algorithms. All the algorithms are implemented in C++ with a MATLAB interface and run on a PC with P4 2.6 GHz CPU and 1024 MB RAM. The code of our algorithm can be downloaded at http://www.comp.polyu.edu.hk/~cslzhang/LBP_Tracking.htm.

The first experiment is on a video sequence of table tennis playing with 58 frames of spatial resolution 352×240. The tracking target is the moving head. The target is initialized as a rectangular region of size 29×39. Since there are distinctive color differences between the target (the head of the player) and the background, all the three models, M1, M2 and M3 locate accurately the target being tracked. To save space, we only show the experimental results by the proposed method in Fig. 4.

Figure 5 illustrates the converging processes of the three methods on frame 2 in the table tennis playing sequence. The iteration numbers of mean shift tracking with M1, M2 and M3 are 5, 4 and 2, respectively. Because M3 suppresses efficiently

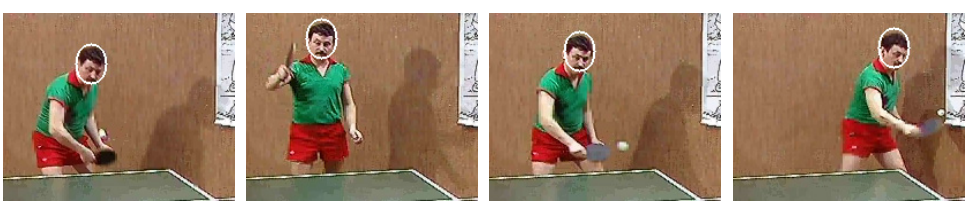


Fig. 4. Tracking results of sequence "table tennis playing" using the proposed method M3. Frames 1, 20, 40 and 58 are displayed.

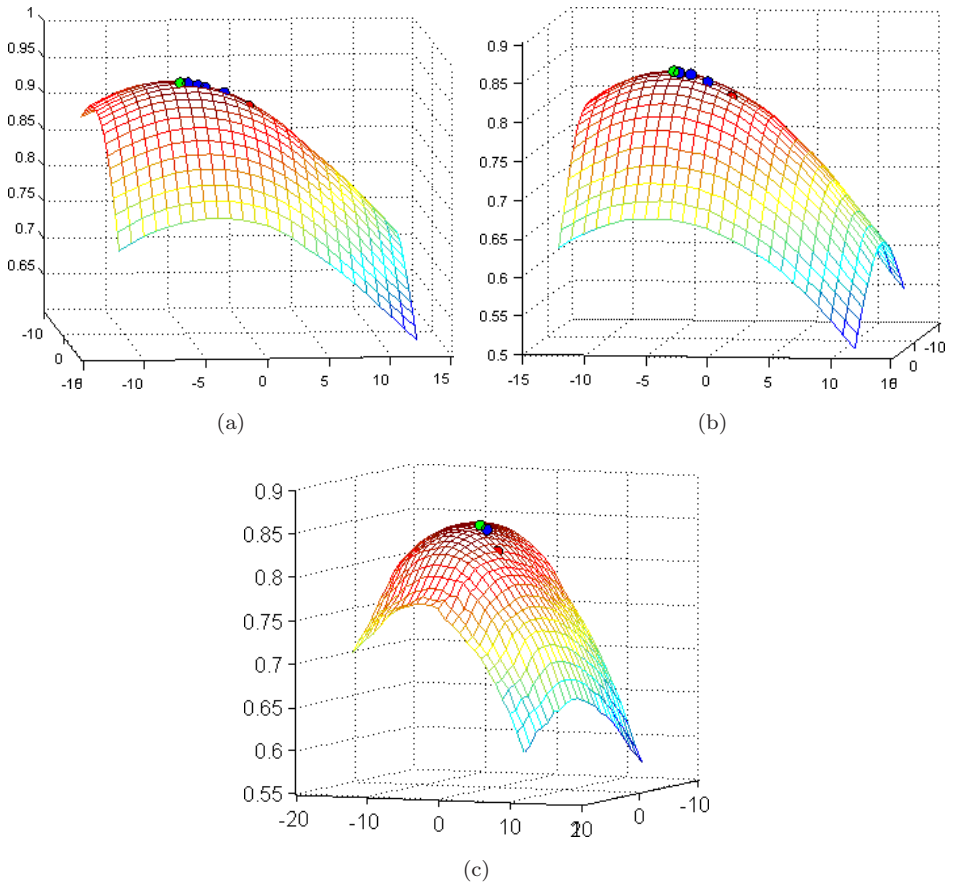
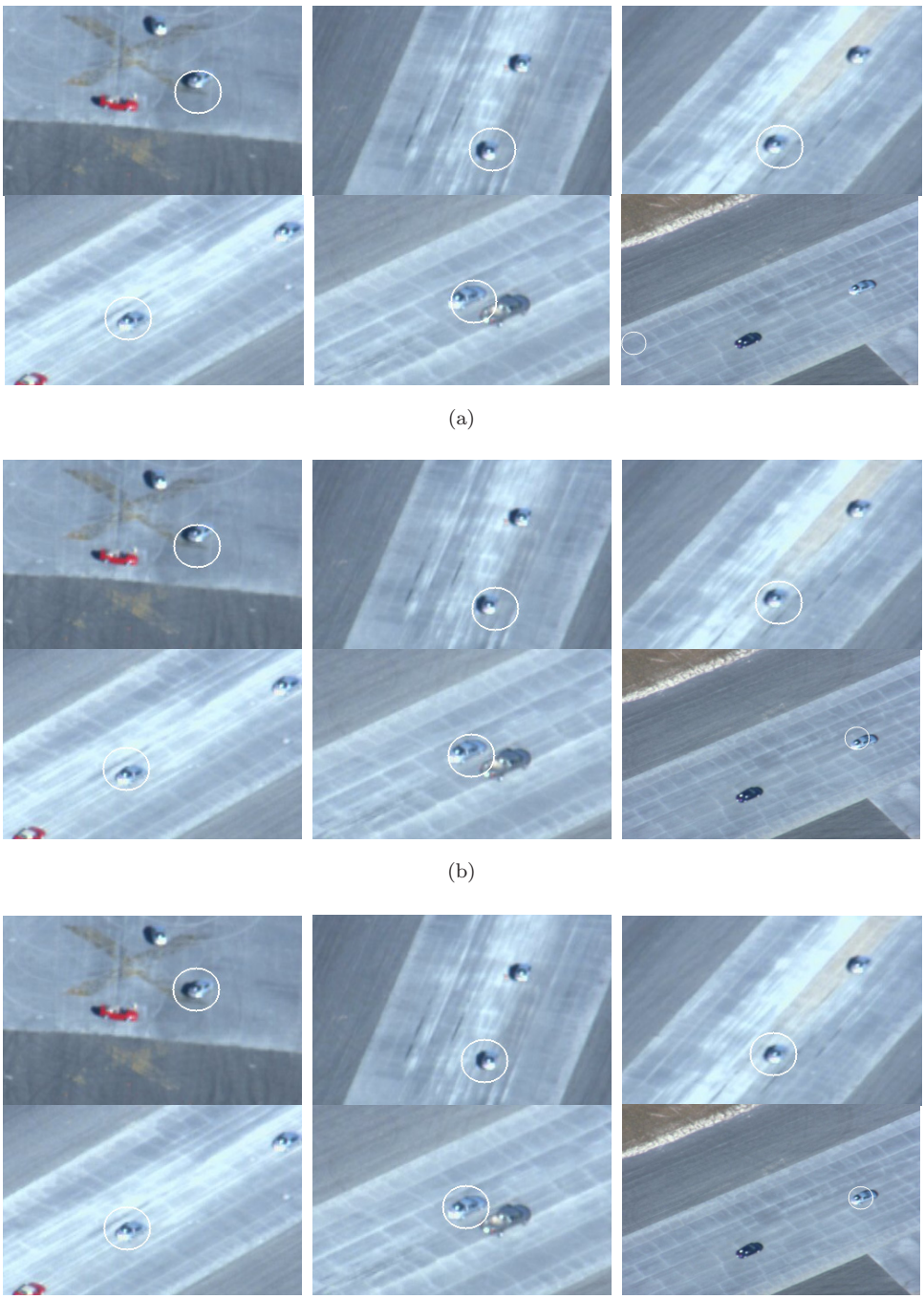


Fig. 5. The converging processes of the three methods on frame 2 in the table tennis playing sequence: (a) M1, (b) M2, and (c) M3. The red dot in the similarity surface represents the initial point, the blue dots represent the points in the iteration process, and the green dot represents the final position. We see that M3 converges much faster than M1 and M2.

the background features, the mean shift tracking with it converges much faster than those with M1 and M2. As can be seen in Table 2, the number of mean shift iterations by the proposed approach is only 54.2% of that by the RGB-based approach.

The second experiment is on a long sequence with 740 frames. In this video, we will track a moving car and compare the target locating accuracies and tracking speed by the three target representations M1, M2 and M3. As shown in Fig. 6, in this scene the target appearance is somewhat similar to the background. Figures 6(a)–6(c) illustrate the mean shift tracking results by the three target representation models. Because the initial target region contains much of the background, the accuracies of the target location of mean shift tracking with M1 and M2 are not good. Especially, mean shift tracking with M1 even loses the object after



(a)

(b)

(c)

Fig. 6. Tracking results of car sequence by the target representation models (a) M1, (b) M2 and (c) M3. Frame 10, 100, 200, 400, 563 and 740 are displayed.

frame 575 due to the interference from another car. However, the proposed model M3 extracts the main target features while suppressing the background features so that the target location is much more accurate than that by M1 or M2 model.

Since the ground truth of the target location in this sequence is available, the localization errors of the three methods are calculated. Table 1 lists the mean and standard deviation of the localization errors by the three methods M1,[a] M2, and M3. The tracking speed is also listed. The experimental results show that the proposed method tracks the car more reliably and more accurately than M1 and M2. It achieves the smallest mean and standard deviation values among the three target representations. In addition, the proposed method tracks 120 frames per second, whose speed is approximately the same as that of the standard mean shift tracking. This is mainly because the proposed method only tracks the key feature points of candidate region with less iterations (refer to Table 2), though it needs additional

Table 1. The target localization accuracies (mean error and standard deviation) and tracking speed by the three methods on the car sequence.

Method	M1	M2	M3
Mean error	8.22	10.78	2.83
Standard deviation	9.15	12.66	3.19
Tracking speed (frames/second)	125	72	120

Table 2. The numbers of mean shift iterations by the three methods.

			Mean Shift Iteration	
Video Sequence	Frames	Target Representation	Total Number of Iteration	Average Number of Iteration
Table tennis	58	M1	286	5.02
		M2	160	2.81
		M3	155	2.72
Car sequence	740	M1	3181	4.30
		M2	2537	3.43
		M3	1612	2.18
Football_1	117	M1	355	3.06
		M2	298	2.57
		M3	253	2.18
Football_2	107	M1	449	4.24
		M2	404	3.81
		M3	255	2.41
Ice skating	300	M1	1456	4.87
		M2	1134	3.79
		M3	1120	3.75

[a]Since the target is lost after frame 575 for M1, we only use the first 575 frames in the calculation for M1.

computation to calculate the LBP features. Overall the proposed method meets the real time requirement.

The third experiment is on a football sequence (football_1), where the task is to track a certain football player. In this sequence, the color of the sports shirt (green) of the target player is very similar to that of the lawn and thus the target and background are hard to distinguish by using only color features. As shown in Fig. 7(a), the M1 model loses the object very quickly. Figure 7(b) shows that the M2 model is slightly better than M1 but it also fails to track quickly. However, Fig. 7(c) indicates that the proposed method successfully tracks the player over the whole sequence.

Fig. 7. Tracking results of the football sequence 1 by the target representation models (a) M1, (b) M2 and (c) M3. Frame 1, 20, 60, 100, 135 and 160 are displayed.

(c)

Fig. 7. (*Continued*)

Figures 8 and 9 show the tracking results on another two sequences. Figure 8 is on a more complex football sequence (football_2). The object player exhibits obvious illumination changes (frames 90 and 107) and partial occlusion (frames 50 and 70). The experimental results show that the proposed method achieves much better performance than M2 and M3, which fail to track quickly. The last experiment is on an ice skating sequence, where the player moves quickly. Figure 9 shows that all the three methods can track the object player; however, the proposed method M3 has much better localization result than M1 and M2.

Table 2 lists the number of mean shift iterations by the three algorithms for the five testing sequences. Compared with M1 and M2, the proposed method M3

(a)

Fig. 8. Tracking results of football sequence 2 by the target representation models (a) M1, (b) M2 and (c) M3. Frame 1, 20, 50, 70, 90 and 107 are displayed.

(b)

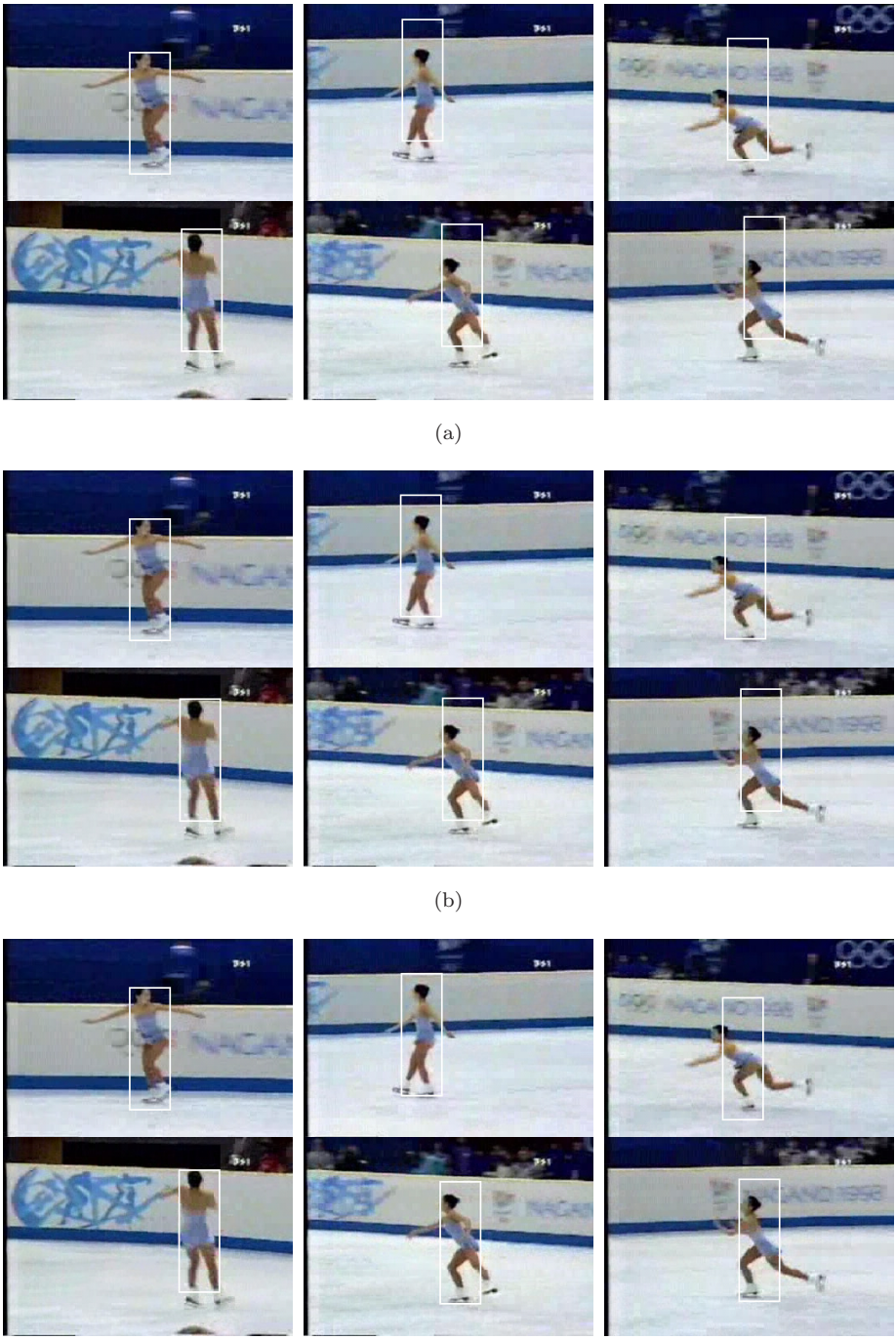
(c)

Fig. 8. (*Continued*)

efficiently extracts the significant feature points of the target and suppresses the smooth background points. Therefore the mean shift tracking with the proposed joint color-texture target representation focuses on the main features in tracking the target, which is very helpful to speed up the convergence of mean shift algorithm. Although the target representation method M2 also decreases the number of iterations, it does not yield good tracking results.

5. Conclusion

LBP operator is an effective tool to measure the spatial structure of local image texture. To reduce the computational cost and improve the robustness of target representation, we proposed a joint color and LBP texture based mean shift tracking algorithm in this paper. A mask of the target is formed based on its five major



(a)

(b)

(c)

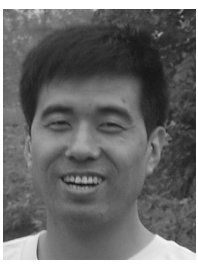
Fig. 9. Tracking results of the ice skating sequence by the target representation models (a) M1, (b) M2 and (c) M3. Frame 1, 60, 120, 160, 260 and 300 are displayed.

uniform $\mathrm{LBP}_{8,1}^{riu2}$ texture patterns and then the target is represented by using its color and texture features within the mask. The proposed target representation model effectively extracts the edges and corners, which are important and robust features, of the object while suppressing the smooth background features. Experimental results indicate that the proposed method performs much better than the original color based method with fewer iteration numbers, especially in tracking objects that have similar color appearance to the background.

References

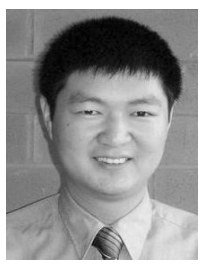
1. A. Adam, E. Rivlin and I. Shimshoni, Robust fragments-based tracking using the integral histogram, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* **I** (2006) 798–805.
2. T. Ahonen, A. Hadid and M. Pietikäinen, Face description with local binary patterns: application to face recognition, *IEEE Trans. Patt. Anal. Mach. Intell.* **28**(12) (2006) 2037–2041.
3. G. Bradski, Computer vision face tracking for use in a perceptual user interface, *Intel Technol. J.* **2**(2) (1998) 12–21.
4. C. Chen and P. Wang, *Handbook of Pattern Recognition and Computer Vision*, 3rd ed. (World Scientific, 2005).
5. Y. Cheng, Mean shift, mode seeking and clustering, *IEEE Trans. Patt. Anal. Mach. Intell.* **17**(8) (1995) 790–799.
6. D. Comaniciu and P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(5) (2002) 603–619.
7. D. Comaniciu, V. Ramesh and P. Meer, Kernel-based object tracking, *IEEE Trans. Patt. Anal. Mach. Intell.* **25**(5) (2003) 564–575.
8. K. Fukunaga and L. D. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Trans. Inform. Th.* **21**(1) (1975) 32–40.
9. C. C. Gotlieb and H. E. Kreyszig, Texture descriptors based on co-occurrence matrices, *Comput. Vis. Graph. Imag. Process.* **51**(1) (1990) 70–86.
10. I. Haritaoglu and M. Flickner, Detection and tracking of shopping groups in stores, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 431–438.
11. M. Heikkiä and M. Pietikäinen, A texture-based method for modeling the background and detecting moving objects, *IEEE Trans. Patt. Anal. Mach. Intell.* **28**(4) (2006) 657–662.
12. V. Jain, A. Ferencz and E. Learned-Miller, Discriminative training of hyper-feature models for object identification, *Proc. British Machine Vision Conf.* (2006), pp. 357–366.
13. Q. A. Nguyen, A. Robles-Kelly and C. Shen, Enhanced kernel-based tracking for monochromatic and thermographic video, *Proc. IEEE Conf. Video and Signal Based Surveillance* (2006), pp. 28–33.
14. K. Nummiaro, E. Koller-Meier and L. V. Gool, An adaptive color-based particle filter, *Imag. Vis. Comput.* **21**(1) (2003) 99–110.
15. T. Ojala and M. Pietikainen, Unsupervised texture segmentation using feature distribution, *Patt. Recogn.* **32**(3) (2000) 477–486.
16. T. Ojala, M. Pietikäinen and T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(7) (2002) 971–987.

17. T. Ojala, K. Valkealahti, E. Oja and M. Pietikäinen, Texture discrimination with multi-dimensional distributions of signed gray level differences, *Patt. Recogn.* **34**(3) (2001) 727–739.
18. M. Pietikäinen, T. Ojala and Z. Xu, Rotation-invariant texture classification using feature distributions, *Patt. Recogn.* **33**(1) (2000) 43–52.
19. M. Pietikäinen and G. Zhao, Local binary pattern descriptors for dynamic texture recognition, *Proc. Int. Conf. Patt. Recogn.* (2006), pp. 211–214.
20. D. Ramanan, D. Forsyth and A. Zisserman, Tracking people by learning their appearance, *IEEE Trans. Patt. Anal. Mach. Intell.* **29**(1) (2007) 65–81.
21. M. Savelonasa, D. Lakovidisa and D. Maroulis, LBP-guided active contours, *Patt. Recogn. Lett.* **29**(9) (2008) 1404–1415.
22. M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Computer Vision*, 3rd ed. (Thomson, 2007).
23. G. Wouwer, P. Scheunders and D. Dyck, Statistical texture characterization from discrete wavelet representations, *IEEE Trans. Imag. Process.* **8**(4) (1999) 592–598.
24. C. M. Wu and K. S. Hsieh, Texture features for classification of ultrasonic liver images, *IEEE Trans. Med. Imag.* **11**(2) (1992) 141–152.
25. C. Yang, D. Ramani and L. Davis, Efficient mean-shift tracking via a new similiarity measure, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* **I** (2005) 176–183.
26. A. Yilmaz, O. Javed and M. Shah, Object tracking: A survey, *ACM Comput. Surv.* **38**(4) (2006).
27. H. Zhang, W. Gao, X. Chen and D. Zhao, Object detection using spatial histogram features, *Imag. Vis. Comput.* **24**(4) (2006) 319–406.
28. G. Zhao and M. Pietikäinen, Dynamic texture recognition using local binary pattern with an application to facial expressions, *IEEE Trans. Patt. Anal. Mach. Intell.* **29**(6) (2007) 915–928.



Jifeng Ning received his Master's degree in engineering from Northwestern A&F University in 2002, and his Ph.D. in 2009 at the State Key Laboratory of Integrated Service Networks, Xidian University. He is now a lecturer in the College of Information Engineering, Northwest A&F University.

His research interests include computer vision, image segmentation and pattern recognition.



Lei Zhang received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, the M.S. and Ph.D degrees in electrical engineering from Northwestern Polytechnical University, Xian, P.R. China, respectively in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, the Hong Kong Polytechnic University. From Jan. 2003 to Jan. 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. Since Jan. 2006, he has been an Assistant Professor in the Dept. of Computing, The Hong Kong Polytechnic University. Dr. Zhang is an Associate Editor of *IEEE Trans. on SMC-C.*

His research interests include image and video processing, biometrics, pattern recognition, computer vision, multisensor data fusion and optimal estimation theory, etc.

David Zhang graduated in computer science from Peking University in 1974 and received his M.Sc. and Ph.D. degrees in computer science and engineering from the Harbin Institute of Technology (HIT), Harbin, P.R. China, in 1983 and 1985, respectively. He received the second Ph.D. degree in electrical and computer engineering at the University of Waterloo, Canada, in 1994. From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Beijing, China, and became an Associate Professor at Academia Sinica, Beijing, China. Currently, he is a Professor with the Hong Kong Polytechnic University, Hong Kong. He is Founder and Director of Biometrics Research Centers supported by the Government of Hong Kong SAR (UGC/CRC). He is also Founder and Editor-in-Chief of the *International Journal of Image and Graphics (IJIG)*, Book Editor of the *Kluwer International Series on Biometrics*, and an Associate Editor of several international journals.

As a principal investigator, he has completed many biometrics projects since 1980. So far, he has published over 200 papers and 10 books.

His research interests include automated biometrics-based authentication, pattern recognition, biometric technology and systems.

Chengke Wu received his B.Sc in wireless communication at Xidian University in 1961. He is a professor with the School of Telecommunications Engineering and the State Key Laboratory of Integrated Service Networks, Xidian University. He was a visiting scholar in the University of Pennsylvania, USA, from 1980 to 1982, visiting professor in Nancy University, France, from 1990 to 1991, and visiting professor in The Chinese University of Hong Kong in 2000, 2001 and 2002, respectively.

As a principle investigator, Prof. Wu has successfully completed many projects, including the 863 High Technology Program of China and Natural Science Foundation of China (NSFC) Key Grant. He has won many awards, published four monographs and over 100 technical papers.

His research interests include image/video coding and transmission, multimedia, computer vision, etc.