Video Anomaly Detection Based on Mixed Statistic Feature

Hao Yu[1], Xinghao Jiang[1, 3], Tanfeng Sun[1, 2, 3], and Shilin Wang[1, 3]
[1]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
[2]State Key Laboratory of Software Engineering, Wuhan University, China
[3]National Engineering Lab on Information Content Analysis Techniques, GT036001, Shanghai, China
Email: {ai_3, xhjiang, tfsun, wsl}@sjtu.edu.cn

***Abstract*—In this paper, a new efficient method based on mixed statistic feature is proposed for video anomaly detection in densely crowded scenes. The proposed mixed statistic feature is a hand-designed feature considering both magnitude and phase information of the optical flow block after the preprocessing step which based on the latent consistency information of moving objects in the block. Gaussian Mixture Model (GMM) is employed in our method to establish the appropriate probability model for our block-based feature. Experimental results on the challenging UCSD datasets (Ped1 and Ped2) have shown that our method outperformed four state-of-the-art approaches both in accuracy and efficiency (less than 1s per frame in Matlab environment).**

***Index Terms*—mixed statistic feature, block-based, GMM, anomaly detection, crowded scenes**

I. Introduction

In our daily life, with the emergence of huge number of surveillance video data, it is extremely significant to detect abnormal behavior in video surveillance automatically, accurately and quickly. In crowded scenes, this challenging problem becomes more complicated. Recently, lots of works have been done in this field [1]-[8]. There is still no generally accepted definition of anomaly in computer vision. However, the main and most common method dealing with the problems tries to establish a comprehensive collection of normal pattern and thus, the abnormal behavior will gain an extremely low probability matching the probabilistic model for normal behavior.

Generally speaking, the existing methods for video anomaly detection can be roughly divided into two categories: moving object tracking based and moving pattern detection based methods.

For the object tracking based methods [1]-[5], the main idea is to track each object in the scene and train a trajectory model. Then the object's motions will be regarded as abnormal behavior when the similarity values between the moving trajectories of the object and the exist trajectories in the training model is low. For non-crowded scenes, this kind of methods could achieve good results; however, for crowded scenes; the object tracking process becomes more complicated due to the ubiquitous and irregular occlusion in complex scenarios.

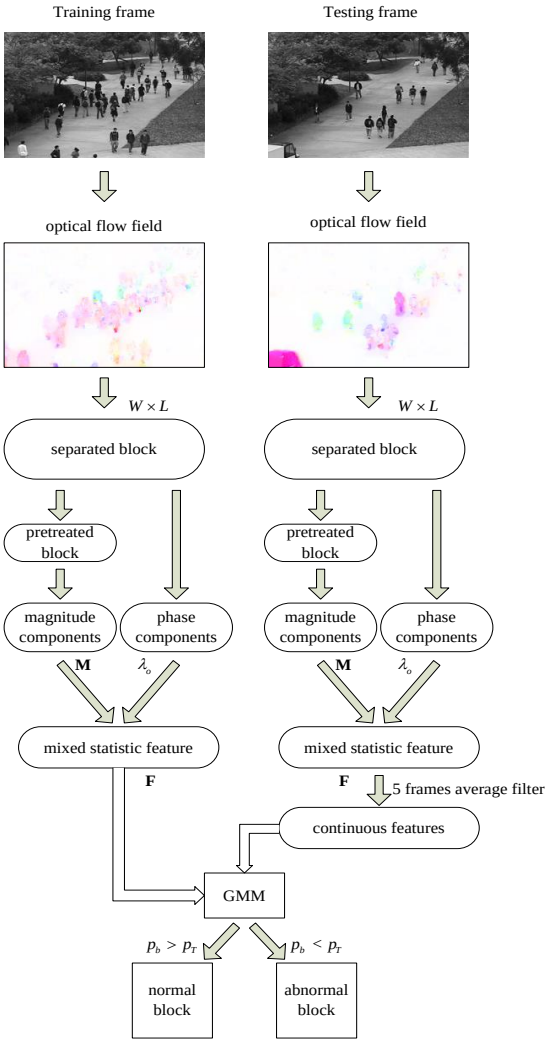


Figure 1. The framework of proposed algorithm

To better detect the abnormal behavior in crowded scenes, the moving pattern detection based methods have been proposed [6]-[8]. Such methods aim to detect the abnormal behavior based on the pixel-level motion features. Contrast to the object tracking based methods, the motion features for a series of pixels or blocks have been analyzed so that the complex object tracking

Manuscript received August 1, 2014; revised November 24, 2014.

doi: 10.12720/ijsps.4.2.150-154

procedure is not necessary for these methods. In [6], Mehran, *et al.* use a social force model based on pure optical flow to detect and localize abnormal behaviors in crowded scenes. Mahadevan, *et al.* propose a framework for anomaly detection and localization in crowded scenes based on mixture dynamic texture model [7]. Xu, *et al*. detect abnormal behaviors in a coarse-to-fine unsupervised learning process based on a hierarchical activity pattern discovery framework [8]. This kind of approaches focus uniquely on the variations of moving patterns based on pixel-level features; however, the consistency information of moving objects is not considered in most pixel-level features for this kind of approaches.

In this paper, we propose a novel and efficient method based on the optical flow. The contributions of this paper are: i) a new hand-designed and mixed statistic feature is proposed to describe the behavior patterns in an efficient manner; ii) a new preprocessing procedure is proposed to eliminate the redundant optical flow information which is irrelevant to the abnormal behavior. This will help us excavate the latent consistency information of moving object. Fig. 1 illustrates the detailed process of our algorithm framework.

II. Video Anomaly Dection

A. Blocks Preprocessing

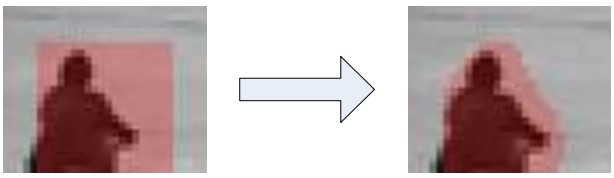
We obtain the whole optical flow field of a frame using the algorithm proposed by Liu, *et al.* [9] in the first step. Then we divide each optical flow field into W×L blocks of size N×N. Obviously, every pixel in the block is actually a vector with two-dimensional information data. Unlike extract a 8-dimension motion feature vector from the block and then group all normalized feature vectors into k clusters in [8], each block has an initialized operating process which includes x-axis and y-axis separated and latent consistency preprocessing of both horizontal direction and vertical direction.

After the coordinates separate process, let $v_x^{(m,n)}$ be the horizontal velocity of the pixel at *m* th row, *n* th column in the block and $v_y^{(m,n)}$ be the vertical velocity of the pixel at *m* th row, *n* th column in the block. Then let $v(x)=[v_x^{(1,1)},...,v_x^{(m,n)},...,v_x^{(N,N)}]\in \mathrm{R}^{N^2}$. And $V(y)=[v_y^{(1,1)},...,v_y^{(m,n)},...,v_y^{(N,N)}]\in \mathrm{R}^{N^2}$ likewise, let. In the latent consistency preprocessing, the $v_x^{(m,n)}$ and the $v_y^{(m,n)}$ will be eliminated from $V(x)$ and $V(y)$ when Inequality (1) and Inequality (2) are true statement at the same time.

$$v_x^{(m,n)} < \mathbf{max}[\mathbf{V}(x)]\times T_v \tag{1}$$

$$v_y^{(m,n)} < \mathbf{max}[\mathbf{V}(y)]\times T_v \tag{2}$$

where max[V(x)] means the maximum value of all elements in V(x), $T_v\in[0,1]$, is a threshold to screen the optical flow of typical region in the block and remove the area that show less consistency with the typical region as shown in Fig. 2. Through this procedure, we combined the advantages of both the object tracking method and the approach based on pixel-level feature. After this preprocessing, we will transform V(x) into V'(x) and transform V(y) into V'(y). The pretreated motion vector V'(x) and V'(y) show stronger consistency both in vector magnitude and vector phase.



original blocks (red) pretreated blocks (red)

Figure 2. The pretreated blocks

B. Mixed Statistic Feature

The mixed statistic feature that we extract from the block consists of two components: the magnitude information and phase information, which are very similar to the sufficient and necessary conditions when we determine a vector. We obtain the magnitude information from the V'(x) and V'(y), which are the result of blocks preprocessing. When $\mathrm{V'(x)}=[v_x^{(1)}, v_x^{(2)}, ..., v_x^{(c)}]\in R^C$ and $\mathrm{V'(y)}=[v_y^{(1)}, v_y^{(2)}, ..., v_y^{(c)}]\in R^C$, where $C\in[1, N^2]$ is the number of pixel selected after the preprocessing. Let $\mathrm{M}=[M_X, M_y]$ be the magnitude components of the block, M_X and M_y can be computed as

$$M_x = \sum_{i=1}^{i=C} v_x^{(i)} \tag{3}$$

$$M_y = \sum_{i=1}^{i=C} v_y^{(i)} \tag{4}$$

To obtain the phase information of the original block, the well-organized coefficient λ_o is proposed. We first calculate the phase of every pixel in the un-pretreated block, and then all the phase will be quantified into 8 region orientations. Let n_i be the number of phase quantified as orientation $i\in\{1,2,3,...,8\}$. Then we define the phase discrete probability distributions vector of the block as $P\in\{p_1,...p_i,...p_8\}\in R^8$, p_i can be computed as

$$p_i = \frac{n_i}{N^2} \tag{5}$$

The information entropy of **P** can be calculated by

$$Entropy = \sum_{i=1}^{i=8}[-p(x_i)\log_2 p(x_i)] \tag{6}$$

Information entropy is an important statistics variable which could evaluates the disordered degree of complex system efficiently. That is exactly contrary to the well-

organized coefficient. Due to the 8 region orientations quantization, the $Entropy \in [0, 3]$. So we could define well-organized coefficient as $\lambda_0 \in [0, 1]$

$$\lambda_o = 1 - \frac{1}{3} Entropy = 1 - \frac{1}{3} \sum_{i=1}^{i=8} [-p(x_i) \log_2 p(x_i)] \quad (7)$$

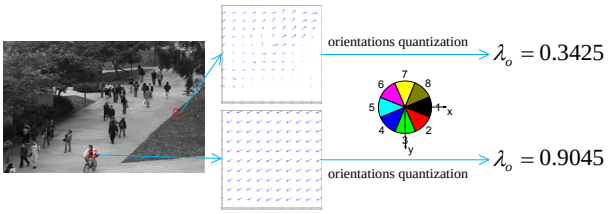


Figure 3. The well-organized coefficient λ_0

The well-organized coefficient λ_0 (as shown in Fig. 3) indicates the phase information while the magnitude information can be represented by M, which is the accumulation of V'(y) and V'(x). Hence, the evaluation of abnormal degree for the whole block can be estimated by mixed statistic feature.

$$\mathbf{F} = \lambda_o \bullet \mathbf{M} \quad (8)$$

C. Gaussian Mixture Model

Gaussian Mixture Model, which has been applied in speaker verification [10] and background subtraction [11] successfully, is employed in this paper to establish the probability model for mixed statistic feature F. The iterative Expectation-Maximization (EM) algorithm will be used to estimate the mean vector μ_κ and covariance matrix Σ_k of Gaussian function from training video sequences.

The probability density function of GMM is given by

$$p(x) = \sum_{k=1}^{K} p(k) p(x \mid k) = \sum_{k=1}^{K} \pi_k N(x \mid \mu_k, \sum_k) \quad (9)$$

where $N(x|\mu_\kappa, \Sigma_k)$ means the posteriori probability.

In the procedure of anomaly detection in testing samples, the mixed statistic feature F is first processed with the 5 frames average filter before fed into the GMM probability model due to the uniformly continuity of both normal behavior and abnormal behavior. At the same time it also harmony with the preference of continuous features for GMM model. Through adopting the method above, every block in each frame of testing video sequences will obtain their own normal probability p_b, then the block will be determined to an anomaly area when p_b is less than the artificial setting threshold p_r.

III. Experimental Result

A. Experimental Setup

The anomaly detection and localization experiments are performed in challenging UCSD Anomaly Detection Dataset for evaluate the accuracy and efficiency of the proposed method. All abnormal behaviors in this dataset, which was split into 2 subsets named Ped1 and Ped2 with different scene, are naturally occurring, rather than manufactured. Specially, both Ped1 and Ped2 are crowded scenes with a variety of anomalies include: runners, wheelchair, skateboarders, bikers, motor vehicles, cart and walking on the lawn. The Ped1 dataset contains 34 training video sequences and 36 testing video sequences and each video frame with a resolution of 238×158. The Ped2 dataset contains 16 training video sequences and 12 testing video sequences and each video frame with a resolution of 360×240.

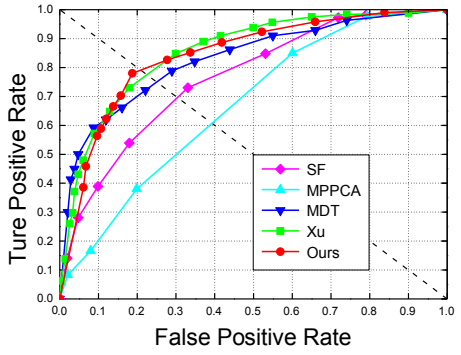
The block size in Ped1 is set to 10×10 and in Ped2 is set to 15×15, which results in each frame consists of 16×24 blocks both in Ped1 and in Ped2. The typical region threshold T_v is set to 0.65 in Ped1and is set to 0.8 in Ped2. The number of Gaussian Model k in GMM is set to 4 both in Ped1 and Ped2.

TABLE I. The AUC on UCSD.

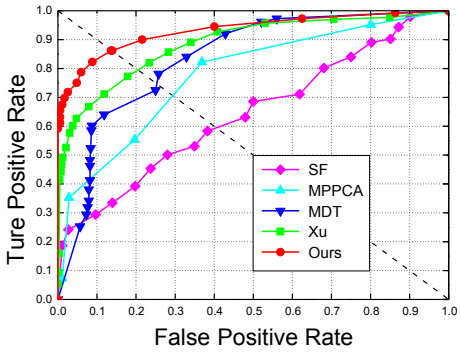
Method	SF [6]	MPPCA [12]	MDT [7]	Xu [8]	Ours
Ped1	67.5%	59.0%	81.8%	**85.4%**	84.2%
Ped2	55.6%	69.3%	82.9%	88.2%	**93.8%**
Average	61.6%	64.2%	82.4%	86.8%	**89.0%**

TABLE II. The EER on UCSD.

Method	SF [6]	MPPCA [12]	MDT [7]	Xu [8]	Ours
Ped1	31%	40%	25%	22%	**21.8%**
Ped2	42%	30%	25%	21%	**13.2%**
Average	36.5%	35%	25%	21.5%	**17.5%**



(a) ROC curve of Ped1 dataset



(b) ROC curve of Ped2 dataset

Figure 4. Evaluation of anomaly detection

B. Accuracy Evaluation

We compared our approach to four other methods: the social force model (denoted SF) [6], the mixture of optical flow (denoted MPPCA) [12], the mixture dynamic texture (denoted MDT) [7] and the method proposed by Xu, *et al.* (denoted Xu) [8].

In Ped1 dataset, as it shown in Fig. 4(a) and Table I, the AUC (Area under Curve) results of our approach is competitive with Xu, the best one of four approaches for comparison. Table II shows our algorithm outperforms all other four methods in the evaluation of EER (Equal Error Rate). In Ped2 dataset, as it shown in Fig. 4(b) and Table I, the method proposed in this paper is vastly superior to the all other four approaches in the evaluation of both AUC and EER. As shown in Fig. 4, Fig. 5 and Fig. 6, the method proposed achieves not only excellent frame-level anomaly detection performance, but also precise anomaly localization on the visual effects in two datasets.

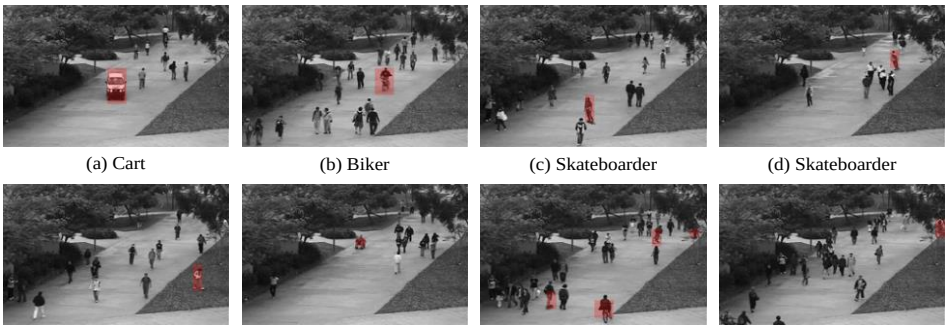
C. Efficiency Evaluation

The processing time for the Xu method and the MDT method is 5s per frame and 25s per frame. In [13], Li, *et al.* improved the MDT method which has been proposed in [7], but it did not present the precise processing time per frame and therefore it seems hard to compare with [13] in efficiency evaluation. Our algorithm under a desktop with 3GHz CPU and 2G memory in the Matlab environment and takes less than 1 seconds per frame in Ped1 dataset while the processing time for obtain the whole optical flow field of a frame needs about 0.8s. That is to say, the efficiency of our method is largely determined by the computation speed of optical flow. So we can improve our computation complexity by only calculate the large magnitude in the optical flow field, rather than the whole field. The code of our method could also be speed up by parallel computing due to the independent calculation of mixed statistic feature for each block.

D. Discussion

It is obvious that the proposed approach performs better in Ped2 than in Ped1 because of the incomplete motion information of two-dimensional image in Ped1 scenes. However, in Ped2, almost all pedestrian movement parallel to the camera plane, which provided accurate relative motion velocity calculate by optical flow.

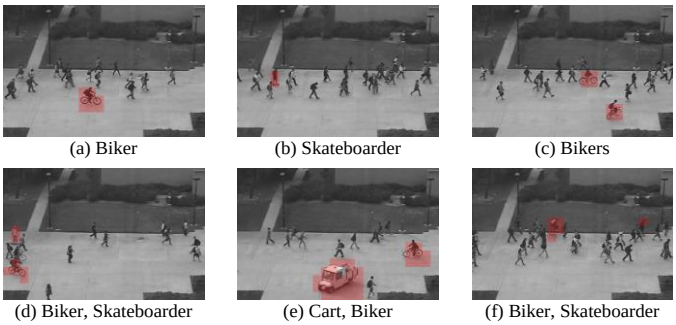
Our algorithm is sensitive to the motion pattern even if there exists irregular occlusion as illustrated in Fig. 3(h), an obstructed biker is in high-speed, which is an anomaly frame obviously, but is labeled as normal frame in the ground truth data provided by Mahadevan, *et al.* [7]. Eliminating these controversial frames, the proposed method could achieve better performs in the evaluation of AUC and EER in Ped1 dataset.



(a) Cart (b) Biker (c) Skateboarder (d) Skateboarder

(e) Crossing-lawn Walker (f) Wheelchair (g) Skateboarder, Bikers (h) Obstructed Biker

Figure 5. Evaluation of anomaly localization in Ped1 dataset



(a) Biker (b) Skateboarder (c) Bikers

(d) Biker, Skateboarder (e) Cart, Biker (f) Biker, Skateboarder

Figure 6. Evaluation of anomaly localization in Ped2 dataset

IV. Conclusion

In this paper, we have proposed a fast method based on mixed statistic feature for anomaly detection and localization in densely crowded scenes. First we calculate the hand-designed feature of each block after obtain the whole optical flow field of all training frames and divide them into blocks. Then GMM probability model will be trained by fed block-based feature into it. Finally, the procedure of anomaly detection in testing frames considering both the preference of continuous features for GMM model and the uniformly continuity of human behavior. Experimental results on challenging UCSD

Anomaly Detection Dataset have shown that our method outperformed four state-of-the-art approaches and could achieve good performance both in anomaly detection and anomaly localization.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (No. 61272439, 61272249), the Fund of State Key Laboratory of Software Engineering, Wuhan University (No. SKLSE2012-09-12) and the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20120073110053).

REFERENCES

[1] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747-757, 2000.
[2] B. T. Morris and M. M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 425-437, 2008.
[3] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
[4] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Video anomaly detection in spatiotemporal context," in *Proc. International Conference on Image Processing*, 2010, pp. 705-708.
[5] T. Zhang, H. Lu, and S. Z. Li, "Learning semantic scene models by object classification and trajectory clustering," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1940-1947.
[6] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 935-942.
[7] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1975-1981.
[8] D. Xu, X. Wu, D. Song, N. Li, and Y. L. Chen, "Hierarchical activity discovery within spatio-temporal context for video anomaly detection," in *Proc. International Conference on Image Processing*, 2013.
[9] C. Liu, *et al.*, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
[10] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.
[11] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. International Conference on Pattern Recognition*, 2004, pp. 28-31.
[12] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2921-2928.
[13] W. X. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18-32, Jan. 2014.

Hao Yu received his B.S. degree in Electrical and Information Engineering from Nangjing University of Posts and Telecommunications, Nanjing, PR China in 2011. He is now a master student of Shanghai Jiao Tong University, Shanghai, China. His current research interests include pattern recognition, computer vision, image and video processing and machine learning.

Xinghao Jiang received the Ph.D. degree in Electronic Science and Technology from Zhejiang University, Hangzhou, PR China in 2003. He is a professor at the School of Information Security Engineering at Shanghai Jiao Tong University, Shanghai, PR China. His current research interests include multimedia security and image retrieval, intelligent information processing, cyber information security, information hiding and watermarking. Dr. Jiang is an IEEE member.

Tanfeng Sun was born in Jilin Province of P.R.China at June 9th 1975. He had his Ph.D. degree on information and communication system from Jilin University in Jilin Province of China at the end of 2003. He worked at Information and Communication Specialty as a Post-doctor in Shanghai Jiao Tong University from Nov. 2003 to Nov. 2005. He works at School of Electronic Information and Electrical Engineering in Shanghai Jiao Tong University in China as an Assistance Professor from 2005 to 2012. He promoted the title of associate professor in Shanghai Jiao Tong University at Dec. 2012 and he still works as a faculty of SJTU now. He had been working at Department of Electrical and Computer Engineering in New Jersey Institute of Technology as a Visiting Scholar from July 2012 to Dec. 2013. His mainly works at publishing papers as follow:

X. H. Jiang, W. Wang, T. F. Sun, Y. Q. Shi, and S. L. Wang, "Detection of double compression in MPEG-4 videos based on Markov statistics," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 447-450, 2013.

T. F. Sun, X. H. Jiang, C. M. Jiang, and Y. Q. Li, "A video content classification algorithm applying to human action recognition," *Electronics and Electrical Engineering*, vol. 19, no. 4, pp. 61-64, 2013.

T. F. Sun, X. H. Jiang, Z. G. Lin, and S. L. Wang, "An H.264/AVC video watermarking scheme in VLC domain for content authentication," *China Communication*, vol. 7, no. 6, pp. 30-36, Dec. 2010.

His research interests are mainly on Digital Forensics on Video Forgery, Digital Image and Video Watermarking, and Video's content recognition and understanding. Professor Sun is an IEEE Member from 2011 to 2014. He had been invited by IEEE North Jersey Section Computer Society Seminar to give a talk on Enormous Challenge to Image & Video Tampering Detection in Fairleigh Dickinson University at Nov. 14, 2013.

Shilin Wang received his B.Eng. degree in Electrical and Electronic Engineering from Shanghai Jiaotong University, Shanghai, China in 2001, and his Ph.D. degree in the Department of Computer Engineering and Information Technology, City University of Hong Kong in 2004. Since 2004, he has been with the School of Information Security Engineering, Shanghai Jiaotong University, where he is currently an Associate Professor. His research interests include image processing and pattern recognition. His biography is listed in Marquis Who's Who in Science and Engineering.