International Journal of Hybrid Information Technology
Vol.8, No.7 (2015), pp.267-276
http://dx.doi.org/10.14257/ijhit.2015.8.7.25

Multi-target Tracking Algorithm Based on TLD under Dynamic Background

Lixia Xue [1], Zuocheng Wang [2] and Yanxiang Chen [1]

1. School computer and information, HeFei University of Technology, HeFei, 230009, China
2. AnHui Sun Create electronic CO., LTD. HeFei, 230088, China
51003239@qq.com, cswangzc@163.com, chenyx@hfut.edu.cn

Abstract

The tracking of multi-target under dynamic background is a challenging issue in computer vision. We improve TLD algorithm by introducing multi-threading mechanism to expand the number of the tracking targets. Thus the excellent framework of TLD for accurate long-time single target tracking is kept, and the tracking of multiple moving targets is realized at the same time. The time cost of tracking also reduces effectively by this algorithm to meet the real-time demand in many dynamic scenes.

Keywords: *TLD, multi-target tracking, multi-threading, real-time*

1. Introduction

Multi-target detection and tracking under dynamic background has been widely used in precision guided weapon, traffic monitoring, visual navigation of mobile robot, intelligent vehicle and unmanned vehicle, etc. As most of dynamic scenes are associated with the target movement, the information of moving targets is helpful to understand the dynamic scenes. In practical applications, accuracy, robustness and real-time are the main requirements to meet: target detection under dynamic background is affected by various factors, including illumination change, background movement, two or more objects sticking together, which will lead to deformation and make difference from the target extracted initially. How to update the target template adaptively is the problem to be solved. In the target tracking process, target occlusion will occur and the correlation information between frames in the image sequences should be applied to improve the robustness of tracking. Meanwhile, the contradiction between accuracy and real-time is also the problem to be solved. To accurately detect and track multi-target, low-level discriminative features such as SIFT [1] (Scale-invariant feature transform) and SURF [2] (Speeded up Robust Features) can be used, while the high complexity of computation of them cannot meet the needs of real-time in the dynamic scenes.

The existing multi-target tracking algorithms can be divided into about 4 categories:

- Multi-feature fusion: As single feature was inadequate to describe the target and not stable for tracking, the algorithms based on multi-feature fusion were proposed. Reference [3] combined color, texture and motion information to describe the target. However, the tracking is easy to fail as the result of background change and target deformation. Reference [4] proposed a multi-target tracking algorithm which combined multi-feature fusion and adaptive template. Reference [5] used multi-feature fusion technology for judging the emergence of a new target, and set a particle filter for each target for realizing the tracking of multiple targets. Due to the complexity of calculation, these algorithms are not suitable for real-time tracking.
- Motion information: Motion information of the target, which can be obtained by optical flow, difference method, etc., is an important feature for target detection. Reference [6] mainly studied person tracking by using difference method to get the

ISSN: 1738-9968 IJHIT

motion information of the subjects. But this method is not suitable for target tracking under dynamic background.

- 3D information: Compared with 2D images, 3D images can solve the occlusion between multiple targets more effectively, and some researchers have studied the multi-target tracking algorithm based on 3D space. Reference [7] proposed a multi-camera 3D single person tracking algorithm based on particle filter, taking advantage of the uniqueness of spatial location of the target to track the severely occluded person accurately. Due to 3D models can deal with the problem of target splitting and merging effectively, reference [8] proposed a 3D multi-target tracking algorithm based on Markov chain Monte Carlo (MCMC). But this method does not apply to target tracking under dynamic background.
- Data association [9]: Cox proposed joint probabilistic data association algorithm (JPDA) [10-12], multiple hypothesis filter (MHF) [13-15] and other data association algorithms, but these algorithms are only suitable for simple scenes. Rasmussen and Hager proposed multi-target tracking algorithm based on multi-feature fusion and data association [16], but it could not solve the problems of target merging and splitting. More importantly, with the growth of the number of targets, the calculation of the algorithm based on data association will increase exponentially, even appearing "combinatorial explosion", which makes it unable to meet the need of practical engineering applications.

In all, the above algorithms have various drawbacks for solving the multi-target tracking problem under dynamic background. This paper attempts to propose an algorithm based on Tracking Learning Detection (TLD) [17] framework. TLD is a long-time tracking algorithm for a single target, in which the deformation, partial occlusion and some other problems in the target tracking process can be solved to some extent with online learning mechanism. When the target re-appears in the view of camera under dynamic background, the algorithm should be able to re-detect it and start tracking again. For multi-target, parallel computing will be introduced through multi-threading mechanism to take full advantage of multi-core CPU, and reduce the time cost as much as possible to achieve real-time tracking of multiple targets.

The remainder of this paper is organized as follows. In Section II, the algorithm process is briefly introduced. In Section III, The experimental results of algorithm are described. In Section IV, we conclude with our major contributions and suggestions for the future work.

2. Multi-Target Tracking Algorithm Based On TLD Under Dynamic Background

In this paper, an effective algorithm for multi-target tracking under dynamic background is proposed by introducing multi-threading mechanism to TLD algorithm.

2.1 Algorithm Flow

As the video is processed in frames, in order to reduce the overhead of creating and destructing of the threads, and enhance the efficiency of the process, we use the thread pool. The framework of algorithm is shown in Figure 1, and the steps can be summarized as follows:

1) Obtaining the first frame image of video, and manually marking N targets to be tracked in the image;
2) Creating N TLD objects. Initializing the corresponding detector and training data of PN learning respectively according to the position of the N targets to be tracked;
3) Obtaining the next frame image, and creating N threads. Making sure that each TLD object corresponds to a thread and tracks a target;

4) tracking result will be displayed when the TLD object successfully tracks a target. For each frame, the main thread can deal with the next frame only after all N threads have completed processing;
5) If the video does not finish, the algorithm jumps to step 3; otherwise, the algorithm exits.

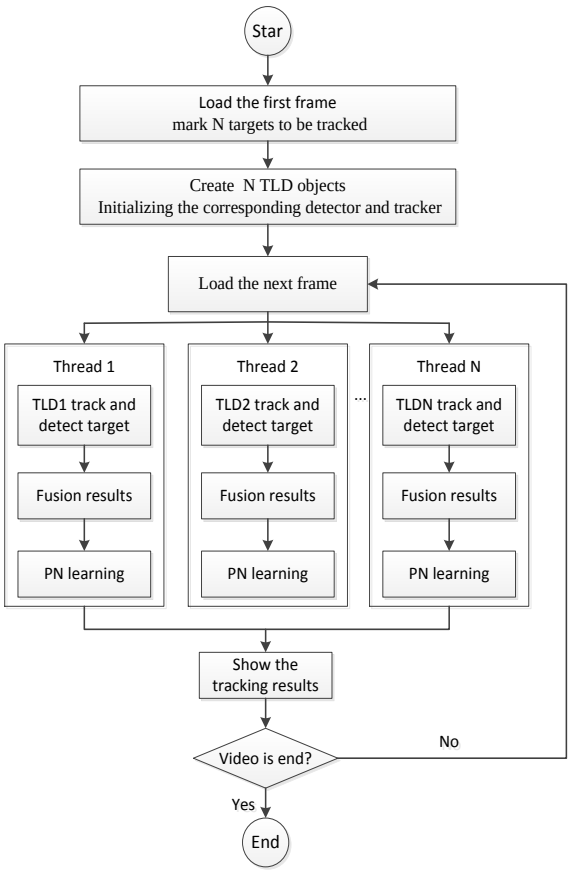


Figure 1. The Framework of the Algorithm

2.2 Target Tracking

First, marking the N targets to be tracked manually, and getting rectangular boxes, then initializing the tracker, detector and initial data of PN learning through the boxes. As in TLD, the tracker is based on pyramid LK optical flow [18] algorithm. Due to the deformation or occlusion, some faults may be produced leading to target lost or tracking error. In order to improve the accuracy of target tracking, we use Media Flow [19] tracking algorithm for tracking calibration.

2.3 Target Detection

The core of target detection is to extract the 2bitBP [20] feature. 2bitBP feature creates four kinds of coding through two comparisons in the extracted area, which has higher efficiency corresponding to the 256 kinds of coding of LBP [21] feature. For each frame the detector scans the image blocks by scanning window, and estimates whether the target is involved in the image blocks. The size of the scanning window is scaled according to the size of the window of the initial target. The detect process has been shown in figure 2, in which the yellow windows are scanning windows.

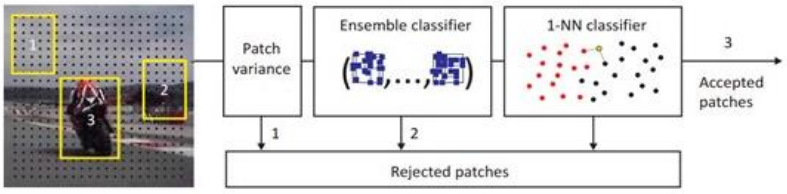


Figure 2. Detector and Detect Process

Each area which may include the target must be judged by variance classification, random forest classifier and nearest neighbor classifier. Any classifier can judge whether it contains the target in the current detection area. Those areas which can be identified to contain the target must be judged by three detectors.

Variance Classifier: The variance of each region is calculated by using the integral image. The region whose variance is less than a threshold would be determined to contain the target.

Random Forest [22] Classifier: The region which has passed the variance classifier is the input of random forest classifier. There are 10 decision trees, each producing a 13-bit binary code x from 13 estimation nodes. Then 10 corresponding posterior probabilities p(y|x) are calculated and the average of them is used to judge that the region would contain the target only if the average is larger than a threshold.

Nearest Neighbor Classifier: When the maximum of similarity is larger than a threshold, the classifier will judge that the region contains the target.

Through target detection the candidate locations of the target can be got, and the most likely location of the target is calculated by fusing the location from target tracking. When the target re-appears in the view of camera under dynamic background, the algorithm is able to re-detect it and start tracking again.

2.4 PN Learning

PN learning [23] can improve the performance of detection gradually. The detection error of the current frame of the video sequence is evaluated to update the classifier, which can avoid the similar error in following frames. PN learning includes four parts: a classifier, training set, supervised learning and P-N experts. The relations between four parts have been shown in Figure 3.

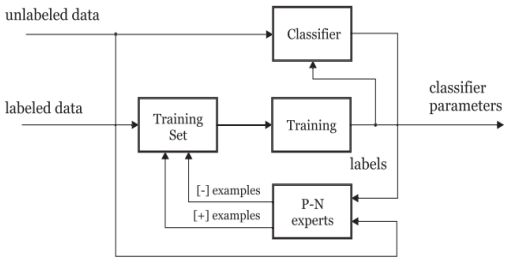


Figure 3. The Flowchart of PN Learning

First, we train an initial classifier by supervised learning based on the labeled data. Then the unlabeled data is classified based on the trained classifier in the previous iteration. The P-N experts are used to identify the misclassified data and correct them, so the performance of the classifier can be improved through the iterative training. The P-expert gives the "positive" label to the data which has been marked as negative by the classifier but should be positive according to the structural constraints. The N-expert

gives the "negative" label to the data which has been marked as positive by the classifier but should be negative according to the structural constraints. This means that P-expert increases the robustness of the classifier and N-expert increases discrimination capability of the classifier.

In the algorithm flow, the detector and the tracker are processed in parallel and complementarily. First, the tracker supposes that the movement of the target between the adjacent frames is limited and the tracked target is visible, so the moving of the target can be estimated. The tracking fails if the target is lost in the video frame. Second, the detector supposes that the frames of the video are independent, and the target is located through global search. Last, PN learning is used to estimate the "positive" and "negative" errors and update the classifier of the detector. Meanwhile, PN learning will also update the "key feature points" of the tracker.

3. Experiment

Experiment is based on Visual Studio2010 development platform. The parameters of the scanning window are as follows：the coefficient of scaling is 1.2, the step size in the horizontal direction is 10% of the width of a video frame, the step size in the vertical direction is 10% of the height of the video frame, the minimum size of the scanning window is 25 pixels.

3.1 The Experimental Results of Algorithm

Scene I: car video. The video resolution is 320 × 240, with a total number of 89 frames. The experimental results have been shown in Figure 4.

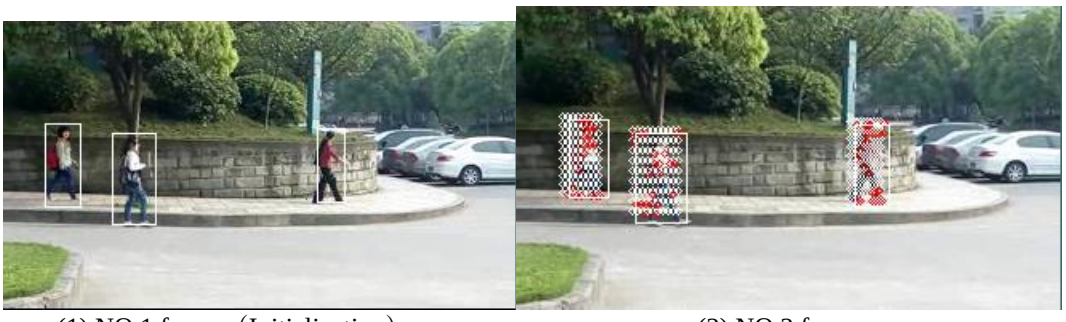
(1) NO.1 frame（Initialization） (2) NO.2 frame

(3) NO.13 Frame (4) NO.35 Frame

(5) NO.64 Frame (6) NO.81 Frame

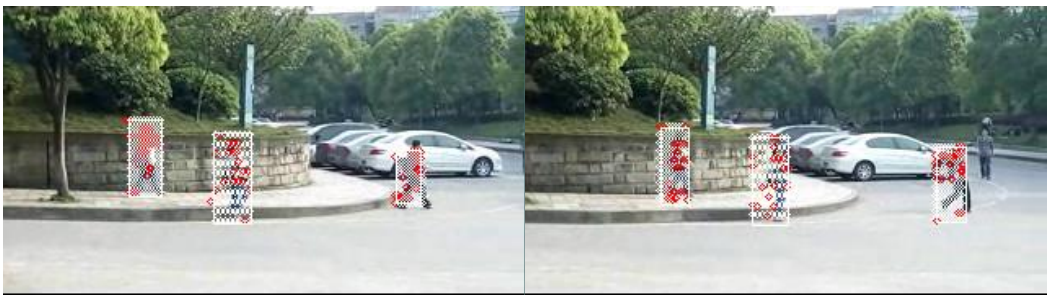
Figure 4. The Results of Multi-Target Tracking In Car Video

Scene II: pedestrian video. The video resolution is 320 × 240, with a total number of 168 frames. The experimental results have been shown in Figure 5.



(1) NO.1 frame（Initialization） (2) NO.2 frame

(3) NO.15 frame (4) NO.43 frame



(5) NO.75 frame (6) NO.117 frame

Figure 5. The Results of Multi-Target Tracking In Pedestrian Video

Only the rectangular boxes in figure 4(1) and figure 5(1) are marked manually. The rest of the rectangular boxes are the tracking results by fusing the results of the tracker and the detector. The white circles are the feature points which are generated by uniform sampling for pixels in rectangular box when tracking. The red circles are the feature points which can be tracked accurately after filtration.

3.2 Test for Occlusion by Background

(1) NO.10frame (2) NO.28 Frame

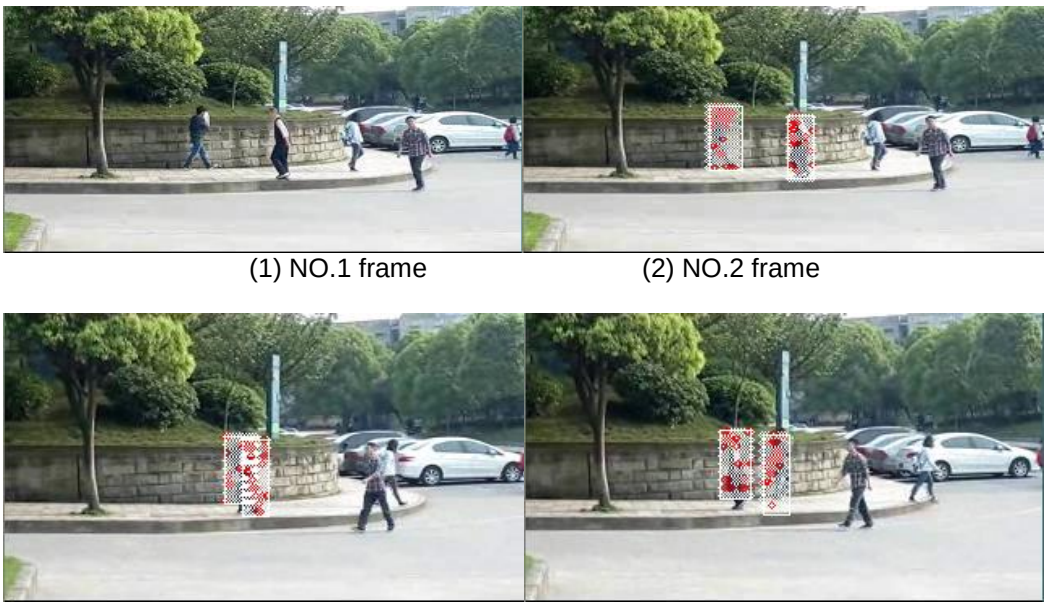
(3) NO.39 frame (4) NO.40 frame

Figure 6. Background Occlusion

In Figure 6, the leftmost person A is the subject we study in this test. In NO.10 frame the occlusion does not occur. In NO.28 frame, person A has been occluded completely by trees. He appears again in NO.39 frame, and the tracker has corrected the tracking position and marked the position of A by tracking box. In NO.40, the tracker has tracked A correctly.

3.3 Test for Occlusion by Targets

Pedestrians2 video clips are used to validate the effect of the algorithm when the targets mutually occlude. The video resolution is 320 × 240, with a total number of 82 frames.



(1) NO.1 frame (2) NO.2 frame

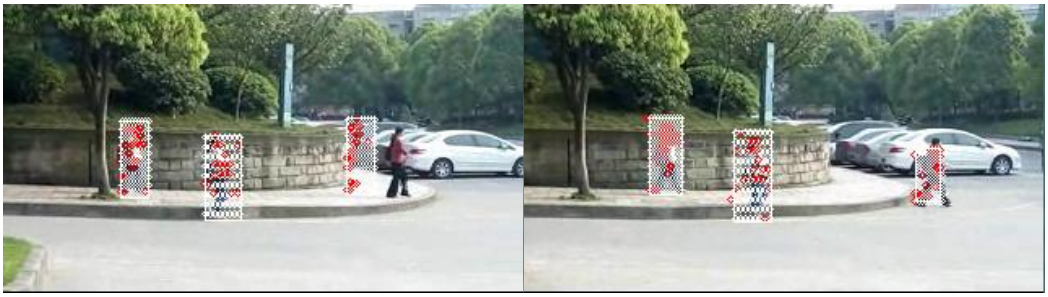
(3) NO.20 frame (4) NO.32 frame

Figure 7. Target Occlusion

In Figure 7, the occlusion does not occur between two persons in NO.2 frame, but it happens in NO.20 frame. In NO.32 frame the two persons separate again. Experimental result shows that the algorithm can track the occluded targets correctly in the tracking process.

3.4 Test for PN Learning Mechanism

In the process of tracking, the tracking box also contains parts of the background. When some corner features of background are relatively clear, the results of tracking will be interfered by these features, even leading to target lost. But PN learning can correct the error in tracking process.



(1) NO.47 frame (2) NO.75 frame

Figure 8. PN Learning Mechanism

In Figure 8, experimental result shows the rightmost people B has been lost in the process of tracking in NO.47 frame because of the interference of the corner features of background. However, the tracker successfully tracks the person B in NO.75 frame because the PN learning mechanism corrects the error.

3.5 Test for real-time

When 1-50 frames of the car video are processed, the time-consuming of the proposed algorithm and TLD for each frame is shown in Figure 9.

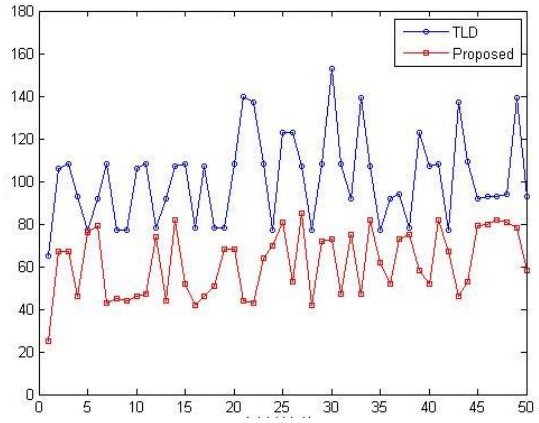


Figure 9. Time-Consuming Comparison

In Figure 9, the blue line and the red line represent the time-consuming of TLD and the algorithm proposed in this paper respectively. Furthermore, the average time of processing a frame by the proposed algorithm is 61ms and TLD algorithm is 101ms. It is obvious that the real-time of the proposed algorithm is better.

4. Conclusion

We propose a multi-target tracking algorithm under dynamic background based on TLD and multi-threading. Not only rigid object like car but also non-rigid object like person can be tracked by the proposed algorithm. Although targets may be lost temporarily during tracking because of the background occlusion or target superposition, PN learning can adjust the results to improve the ability of anti-occluding. In addition, the experiment results show that the multi-threading mechanism can enhance the real-time of tracking.

Acknowledgements

This work is partially supported by the National Nature Science Foundation of China (61105076), China Post-doctoral Science Foundation (2012M511402), and Fundamental Research Funds for the Central Universities of China (JZ2014HGBZ0059).

References

[1] Lowe D G. Distinctive image features from scale-invariant key-points. International journal of computer vision, **2004**, 60(2): 91-110.
[2] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF). Computer vision and image understanding, **2008**, 110(3):346-359.
[3] Takala V, Pietikainen M. Multi-object tracking using color, texture and motion. Computer Vision and Pattern Recognition, **2007**. CVPR2007. IEEE Conference on. IEEE, 2007:1-7.
[4] Kumar P, Brooks M.J. An adaptive bayesian technique for tracking multiple objects. Pattern Recognition, **2007**:657-665.
[5] Pernkopf F. Tracking of multiple targets using online learning for reference model adaptation. IEEE Trans on SMC-B, **2008**, 38(6):1465-1475.
[6] Chang FL, Ma L, Qiao YZ. Human oriented multi-target tracking algorithm in video sequence. Control and Decision, **2007**, 22(4):418-422.
[7] Yao Jian, Odobez JM. Multi-camera 3D person tracking with particle filter in a surveillance environment. EUSIPCO, **2008**:25-29.
[8] Osawa T, Sudo K, Arai H. Monocular 3D tracking of multiple interacting Targets. ICPR, **2008**:1-4.
[9] Cox IJ. A review of statistical data association techniques for motion correspondence. Compute Vision, **1993**, 10(1):53-66.
[10] Messaoudi Z, Ouldali A, Oussalah M. Joint multiple target tracking and classification using controlled based cheap JPDA-multiple model particle filter in cluttered environment. ICISP, **2008**:562-569.
[11] Lee HK, Ko HS. Predictive estimation method to track occluded multiple objects using joint probabilistic data association filter. ICIAR, **2005**:852-860.
[12] Shafique K, Lee MW, Haering NC. A rank constrained continuous formulation of multi-frame multi-target tracking problem. CVPR, **2008**:1-8.
[13] Chia AYS, Huang WM. Multiple objects tracking with multiple hypotheses dynamic updating. ICIP, **2006**:569-572.
[14] Chia AYS, Huang WM, Li LY. Multiple objects tracking with multiple hypotheses graph representation. ICPR, **2006**:638-641.
[15] Betke M, Hirsh DE, Bagchi A.Tracking large variable numbers of objects in clutter. CVPR, **2007**:1-8.
[16] Rasmussen C, Hager GD. Probabilistic data association methods for tracking complex visual objects. IEEE Trans Pattern Anal Mach Intell, **2001**, 23(6):560-576.
[17] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, **2012**, 34(7):1409-1422.
[18] Bouguet J Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation, **2001**, 2:3.
[19] Kalal Z, Mikolajczyk K, Matas J. Forward-backward error: Automatic detection of tracking failures. IEEE International Conference on Pattern Recognition, **2010**:2756-2759.
[20] Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking. IEEE International Conference on Computer Vision Workshops, **2009**:1417-1424.
[21] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, **2002**, 24(7):971-987.
[22] Breiman L. Random forests. Machine learning, **2001**, 45(1):5-32.
[23] Kalal Z, Matas J, Mikolajczyk K. Pn learning: Bootstrapping binary classifiers by structural constraints. IEEE Conference on Computer Vision and Pattern Recognition, **2010**:49-56.

Authors

LiXia Xue, she received the Ph.D. degree from the School of Civil Engineering, Southwest Jiaotong University, Chengdu, China. She worked in College of computer science and technology, Chongqing University of posts and telecommunications from 2002 to 2010. From 2010 to 2013, she worked in postdoctoral workstation of University of science and technology of China and the 38th institute of china electronics technology group from 2010-2013. Since 2013, she working for School of computer and information, Hefei university of technology Hefei, China. Her research interests include digital image processing and pattern recognition.

ZuoCheng Wang, He is an Associate Professor with the Software Institute of Chongqing University of posts and telecommunications, Chongqing, China. He received the Ph.D. degree from the Southwest Jiaotong University in 2007. He worked in postdoctoral workstation of Beijing University and the 38th institute of china electronics technology group from 2008-2010. Since 2010, he working for Anhui sun create electronic CO.LTD.as deputy General Manager. His research interests include public security and smart city.

Yanxiang Chen, She received the Ph.D. degree from the Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, China. She was a Visiting Scholar with the Beckman Institute, University of Illinois at Urbana-Champaign, Champaign, IL, USA, from 2006 to 2008, and with the Department of Electronic Computer Engineering, National University of Singapore, Singapore, from 2012 to 2013. She is an Associate Professor with Hefei University of Technology, Hefei. Her work has been supported by the Natural Science Foundation of China and the National High-Tech Research and Development Program of China. Her research interests include audio-visual signal processing, saliency, and scene analysis..