International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.5 (2016), pp.431-440
http://dx.doi.org/10.14257/ijsip.2016.9.5.38

Image Tracking Algorithm Improvement Based on TLD Frame

Lei Yu, Tao Zheng and Qiyun Shi

College of Information and communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China
Corresponding Author: Lei Yu
E-mail: yulei@hrbeu.edu.cn

Abstract

Image target tracking is an important part in computer vision field, and has a broad application prospect in modern society. Lack of online learning function, the traditional target tracking algorithm was based on the single track as long as possible. The target deformation and occlusion will lead to target tracking lost. TLD (tracking-learning-detection) is the long-term effective algorithm in tracking single target proposed. The algorithm requires less prior knowledge and can achieve long-term tracking, but when target encounter obstacles, there will be drift phenomenon in the short term. In order to overcome this shortcoming, the paper improve the tracking module of TLD. A method is proposed by using Harris feature point instead of the original Grid sampling method. The experimental results show that this method can effectively suppress the tracking drift, and improve tracking precision.

Keywords: *TLD, long-term tracking, Harris, drift suppression*

1. Introduction

Image target tracking is one of the most important research problems in computer vision field, and widely used in the medical, video monitoring, artificial intelligence, military *etc.* typical Tracking method in general can be divided into four categories: (1) tracking based on the feature points (including the sparse tracking method and dense tracking method) . sparse tracking method only track the target points representing the features, such as Lucas-Kanade optical flow method [1]. dense tracking methods track all points of the whole frame image, such as Horn_Schunck, block matching method. compared to the sparse tracking method, computation of dense tracking is obviously much larger. (2) tracking based on estimation, such as mean-shift[2], CAMSHIFT, Kalman[3], Condensation[4], particle filter[5]; (3) tracking based on the template, such as the target object contour tracking; (4) tracking based on detection, such as Adaboost; (5) tracking based on transform, such as Hough transformation. These typical tracking methods can be called short-term tracking algorithm.

In order to solve the shortage of short-term tracking algorithms, scholars began to design long-time target tracking approaches. A successful long-term tracking algorithm has to meet the following requirements: (1) to keep track of the target in a very long video sequence; (2) when the target is occluded, it is still able to capture target, especially when the target appearance changes; (3) can overcome the target scale changes, illumination changes, background interference and other issues. Based on the above considerations, the TLD target tracking algorithm which is proposed by Kalal is a very effective algorithm for single target tracking.

Kalal[6-9] proposed a new detection system that has three parts to finish detection: tracking, learning, and detection (TLD). TLD is a effective method that be able to achieve a single target tracking algorithm in the long term, has strong robustness, and very strong learning ability of recovery.

ISSN: 2005-4254 IJSIP

2 .TLD Method

2.1. TRACKING -LEARNING -DETECTION

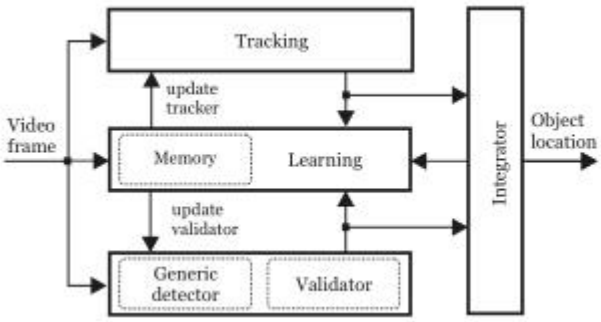


Figure 1. The System Structure of the TLD

TLD is a new detection system designed for tracking object in a video. The structure is shown in Figure 1. Tracker employs LK median value pyramid method to estimate the motion between frames. Detector is cascaded classifier structure, and has three part: variance detector, collection detector and nearest neighbor detector. The learning module is evaluated by P-N learning to evaluate the results of the tracker and the detector, and the results are classified, and the P-N sample database is updated. Integrator module analyze the data from the tracker and the detector. If the results of the tracker data is greater than the detector, the target is output.

As other target detection methods, the detection module in TLD may also has errors, and the error is nothing more than the wrong negative samples and the wrong positive samples. The learning module is based on the results of the tracking module to evaluate these two errors, and generate training samples according to the evaluation results. It can be seen that the tracking module is very important in the whole TLD system, which will direct affect the final results.

2.2. TLD Tracker[9]

TLD tracking module is based on the front and back error of the Pyramid median optical flow tracking method. First of all, take the tracking object feature points sampling, also be referred to as the template frame sampling A grid of 10×10 points is employed to estimate the motion with LK pyramid tracker. LK pyramid tracker uses two levels of the pyramid and represents the points by 10×10 patches. In order to ensure that the feature points are tracked correctly, the TLD introduces Forward-Back-Error and the similarity verification mechanism.

For object image tracking, points tracking is the work to be faced. If the selected feature points disappeared or changed, point tracking is likely to fail.

The right track is: there is a target P0 in the t-th frame, tracking algorithm gets a point P1 in the t+1 th frame, assuming P1 is right tracking point, applies reverse tracking from P1 of the t+1 frame back to P2 of the t-th frame. If P0 and P2 are coincidence, this is the right track.

Fig. 2 shows tracking results of the target point in two consecutive frames. point 1 has a visible prominent feature, and is tracked correct through the forward and backward path verification. because the feature of Point 2 is not obvious and the background is black, the position offset appeared in the tracking process.

Tracking module of TLD obtains FBError values of 10×10 Grid points through the forward and backward tracking, and then take the middle value midFRError as the threshold. Remove the points of the t+1 frame if their FBError value are greater than the threshold, and keep the less. Finally the trace output is determined based on these target points.

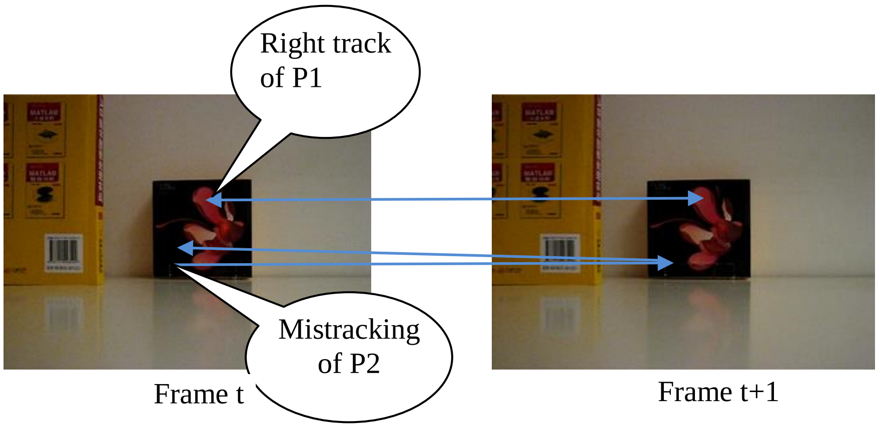


Figure 2. Forward-Back-Error

2.3. Drawbacks of TLD

Obviously, as a long-term image target tracking algorithm, TLD is very effective. It can effectively deal with object occlusion, deformation, illumination changes and occlusions. However, when encountered slow moving, it will produce tracking drift. The situation is illustrated by the following example:

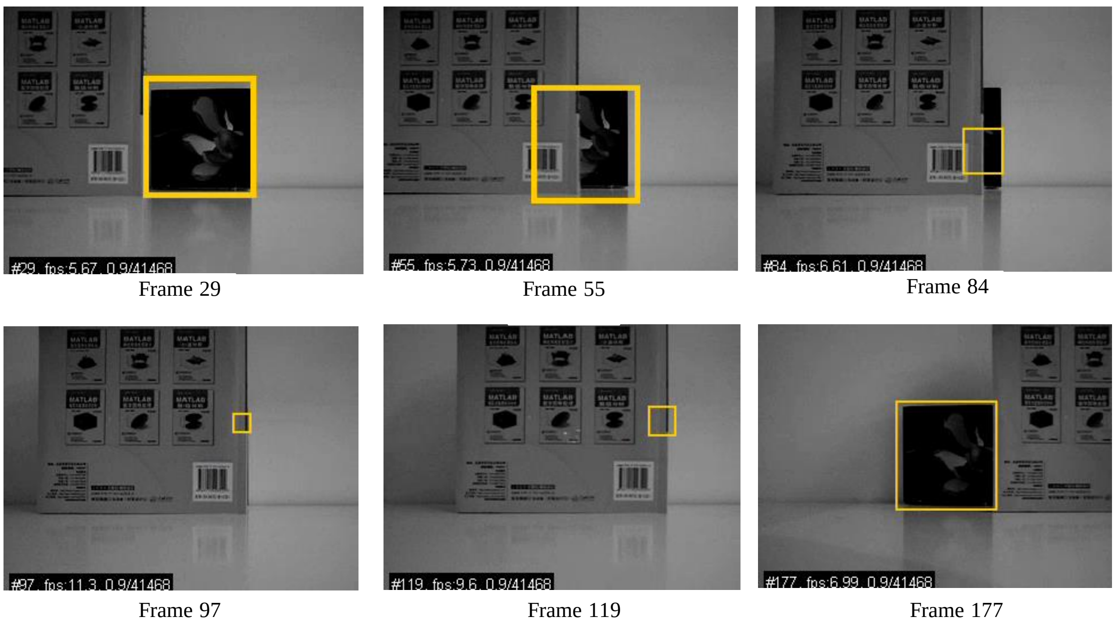


Figure 3. Tracking Results of TLD Algorithm

In Figure 3, the target is a box, and a book slowly moves to block the box from left to right. When the box is sheltered to 1/2, tracking become unstable. From frame 97 to frame 119, the target has been completely obscured, but tracking box still display, tracking result is books edge. tracking has shift. the reasons of this tracking shift are analyzed as follows. the tracker using the Grid uniform mode in tracking features. when the target is different from the background, tracking is generally normal, because mostly sampling points taken by Grid mode can be representative the characteristics of the target. However, when there is a disturbance of similar objects, tracking can appear drift.

Therefore, this paper proposes an improved algorithm employing Harris feature instead of Grid mode.

3. The Improved TLD with Harris Corner

3.1. Harris Corner

Recognition of human diagonal points is usually finished in a small local window. If moving this particular small window in all directions, great changes have taken place in the grays of pixels in the window, then there will be a corner, otherwise the gray scale in the window does not change, there is no corner[10-12].

In a image $I(x,y)$, assume point (x,y) moves $(\Delta x,\Delta y)$. the Self-similarity of move-ment is represented by autocorrelation function.

$$c(x,y,\Delta x,\Delta y)=\sum_{(u,v)\in W(x,y)} w(u,v)(I(u,v)-I(u+\Delta x,v+\Delta y))^2 \tag{1}$$

In Equation(1), $W(x,y)$ is the window whose center point is (x,y). $w(u,v)$ is weighted function, also can be Gauss weighted function. According to the Taylor expansion, a first-order approximation of movement $(\Delta x,\Delta y)$ for image $I(x,y)$:

$$\begin{aligned} I(u+\Delta x,v+\Delta y) &\approx I(u,v)+I_x(u,v)\Delta x+I_y(u,v)\Delta y \\ &= I(u,v)+[I_x(u,v),I_y(u,v)]\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \end{aligned} \tag{2}$$

In Equation(2), I_x、 I_y are the partial derivatives of the image $I(x,y)$. Then Equation(1) is approximated to

$$\begin{aligned} c(x,y;\Delta x,\Delta y) &= \sum_w (I(u,v)-I(u+\Delta x,v+\Delta y))^2 \\ &\approx \sum_w ([I_x(u,v)I_y(u,v)]\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix})^2 \\ &= [\Delta x \quad \Delta y]M(x,y)\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \end{aligned} \tag{3}$$

$$M(x,y)=\sum_w \begin{bmatrix} I_x(u,v)^2 & I_x(u,v)I_y(u,v) \\ I_x(u,v)I_y(u,v) & I_y(u,v)^2 \end{bmatrix} \tag{4}$$

It shows, Equation(1) could be approximated to a binomial function:

$$c(x,y;\Delta x,\Delta y)\approx[\Delta x \quad \Delta y]M(x,y)\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \tag{5}$$

The binomial function is essentially an elliptic function. Oblations and size of an ellipse are determined by feature values λ_1、 λ_2, and elliptical direction is determined by feature vectors of $M(x,y)$.

Identification method of Harris corner does not need to calculate the specific feature values, but a corner response value R. The calculation of R is:

$$R=\det M-\alpha(traceM)^2 \tag{6}$$

In equation(6), $\det M$ is the determinant of Matrix $M(x,y)$. α is a empirical constant , and the range is 0.04 ~ 0.06.

The calculation steps of Harris corner are as follows:

(1) calculating the gradients I_x 、 I_y of image $I(x,y)$ in two directions x and y .

$$I_x = \frac{\partial I}{\partial x} = I \otimes (-1 \quad 0 \quad 1) \tag{7}$$

$$I_y = \frac{\partial I}{\partial y} = I \otimes (-1 \quad 0 \quad 1)^T \tag{8}$$

(2) Calculate the product of two gradients, $I_x^2 = I_x \cdot I_x$, $I_y^2 = I_y \cdot I_y$, $I_{xy}^2 = I_x \cdot I_y$.

(3)Applying Gauss weighted function t $I_x{}^2$ 、 $I_y{}^2$ and I_{xy} , and generating the elements of matrix M .

$$A = g(I_x{}^2) = I_x^2 \otimes w\,,\; B = g(I_y{}^2) = I_y^2 \otimes w\,,\; C = g(I_{xy}) = I_{xy} \otimes w$$

(4) Calculating the Harris response value R of each pixel, and setting R to zero if less than certain threshold.

(5) Maximum Suppression in 3×3 or 5×5 neighborhood, and The local maxima point is Harris corner.

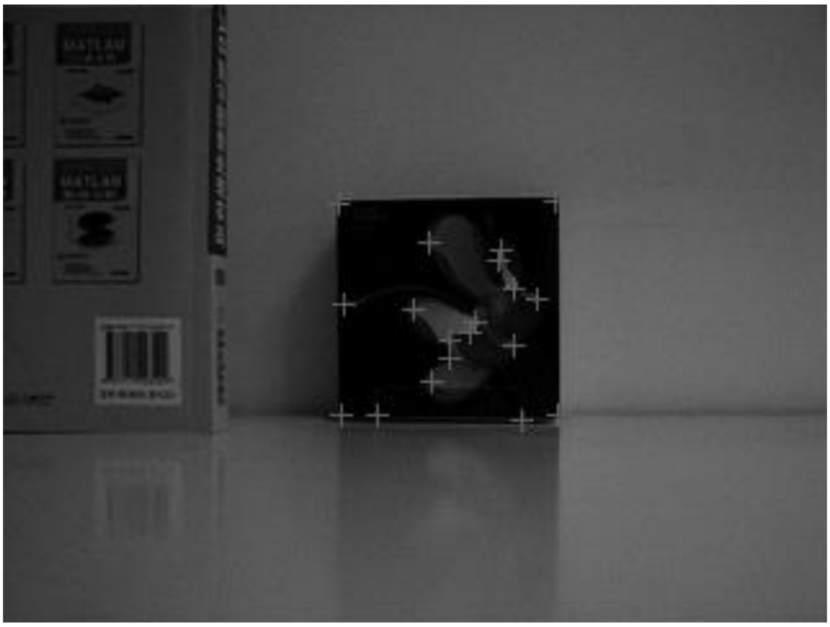


Figure 4. Detection Results of Harris Corners

Figure 4 is the detection results of Harris corners . It shows that the key features of the target box have been detected. these feature points can act as the tracker target. good feature points are very important to the performance of the tracker.

3.2. Analysis of Experimental Results

Software environment: MatlabR2010a, Opencv2.2, visual C++2008; Genuine 1.83GHZ. hardware environment: Intel Genuine 1.83GHZ, memory 1GB. The video resolution is 320*240. There are two experimental videos. Video 1: tracking a walking pedestrian. there will be other pedestrians near the target people as the interference. Video

2: the target is a box, with a book moving slowly to occlude object from left to right. the intermediate process includes: the beginning of occlusion, half occlusion, total occlusion, exposed target.

In Experiment 1, when there are other pedestrians pass through target pedestrians, the tracking by original TLD method appeared obvious drift from frame 77 to frame 79, and relocked the correct target in frame 80, while the improved TLD algorithm did not appear such drift. The results show in the Figure 5 and 6. In Experiment 2, when the book cover half of the box, tracking become unstable from Frame 55. The object was completely covered in Frame 97, however the tracking frame of original TLD method still display, and the tracking result is the books of the edge, as shown in Figure 7 . are shown, until the target box reappear on the screen in Frame 199, the target was relocked. the improved TLD algorithm solved this problem and realize the better tracking, as shown in Figure 8.

Frame 12 Frame 44

Frame 77 Frame 80

Figure 5. Missing Tracking Results of Original TLD

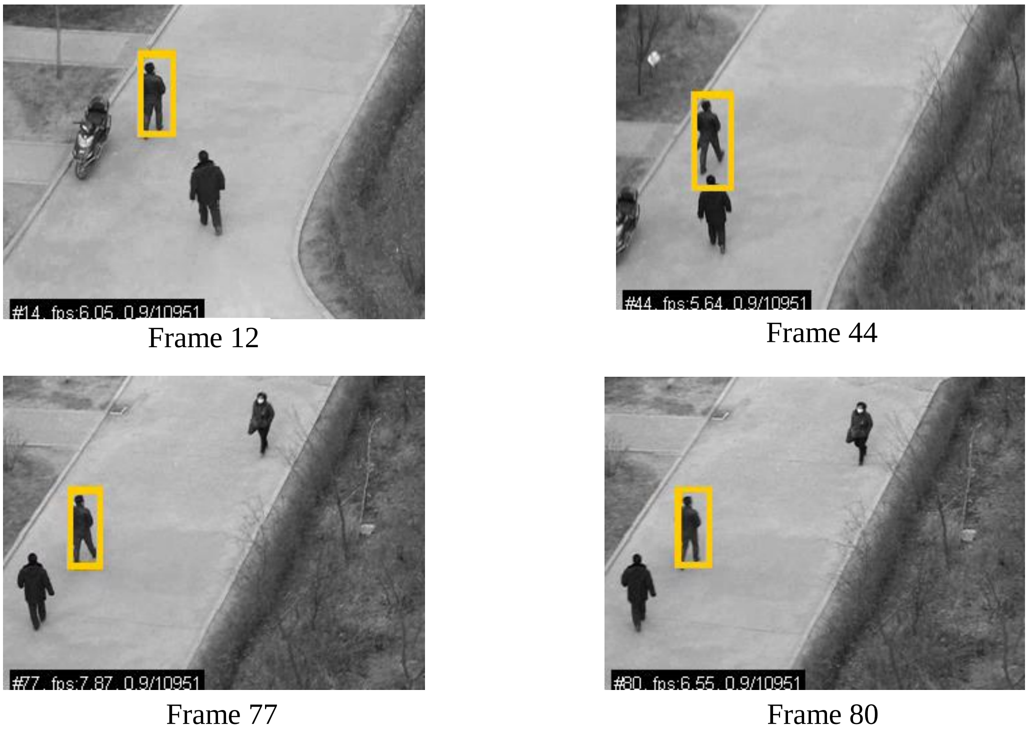


Figure 6. Better Tracking Results of Improved TLD

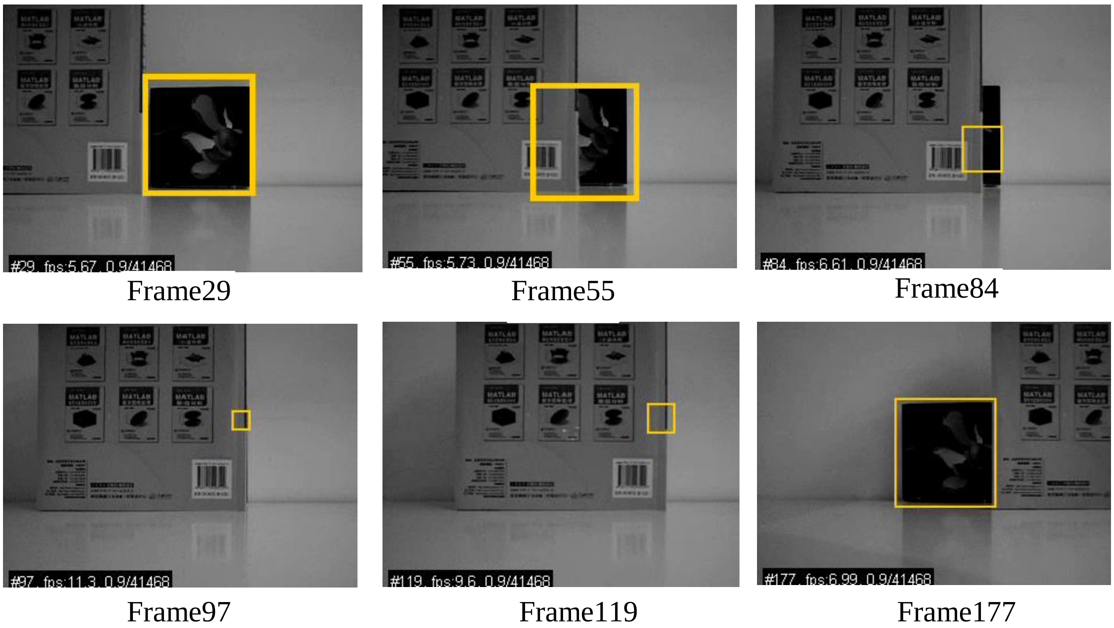
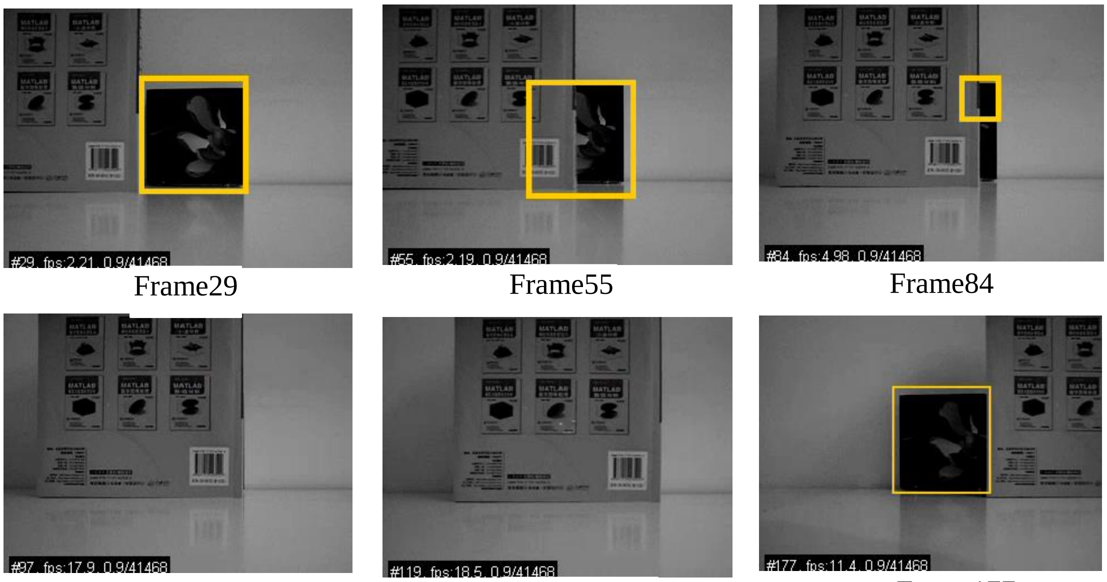


Figure 7. Missing Tracking Results of Original TLD



Frame29 Frame55 Frame84

Frame97 Frame119 Frame177

Figure 8. Better Tracking Results of Improved TLD

4. Conclusion

Object occlusion is a difficult problem for each target tracking algorithm. TLD as a very effective long-time tracking algorithm, overcomes the shortcomings of other algorithms which cannot achieve online learning function, and solved the problem of long-term tracking, but facing similar obstacles, TLD will appear tracking drift in the short term, and will affect the actual tracking effect. The paper proposed an improved TLD method using Harris corner instead of the original Grid sampling point. The experiment results showed it is a good solution to the problem of tracking drift.

Acknowledgements

This paper is supported by Fundamental Research Funds for the Central Universities (GK2080260103), and NSFC(61003128).

Reference

[1] J. L. Barron, D. J. Fleet and S. S. Beauchemin, “Performance of Optical Flow Techniques”, International Journal of Computer Vision, vol.12, no.1, **(1994)**, pp.43–77.

[2] F. G.Bajramovic and J. C. Denzler, “Efficient combination of histograms for real-time tracking using mean-shift and trust-region optimization”, Austria: Springer Verlag, Heidelberg, **(2005).**

[3] D.Terzopoulos and R. Szeliski, “Tracking with Kalman Snakes”, Active Vision. MIT Press, Cambridge, Massachusetts, **(1992)**

[4] M. Isard and A. Blake, “Condensation–Conditional Density Propagation for Visual Tracking”, International Journal of Computer Vision, vol. 29, no.1, **(1998)**, pp.5-28

[5] L. Shuai and L. Ke-zheng, “Scale-invariant Feature Transform Based on Clustering in Face Recognition”, Journal Of Harbin University Of Science And Technology, no.1, **(2014)**,pp. 31-35

[6] Z. Kalal, J. Matas and K. Mikolajczyk, “Online learning of robust object detectors during unstable tracking”, Proceedings of the 3rd On-line Learning for Computer Vision Workshop, Kyoto Japan**, (2009)**, Sept. 27-Oct. 4.

[7] Z. Kalal, K. Mikolajczyk and J. Matas, “Forward-backward error: automatic detection of tracking failures”, Proceedings of International Conference on Pattern Recognition, Istambul Turkey **(2010)**, Aug.23-26.

[8] Z. Kalal, K. Mikolajczyk and J. Matas, “P-N learning: bootstrapping binary classifiers by structural

constraints", Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition San Francisco ,USA **(2010)**, June13-18;

[9] Z. Kalal, K. Mikolajczyk and J. Matas, "Tracking-Learning-Detection . IEEE Transactions on Pattern Analysis and Machine Intelligence, , vol. 34, no. 7, **(2012)**, pp.1409 – 1422.

[10] Z. Xin, Q. Qiumeng, Y. Yongqiang and W. Congqing, "Improved TLD visual target tracking algorithm", Journal of Image and Graphics, vol. 18, no.9, **(2013)**, pp.1115-1123.

[11] S. R. Mokhtarianf, "Curvature scale space based image corner detection", Proc． European Signal Processing Con, Island of Rhodes, Greece, **(1998)**, Sep. 5-8.

[12] L.Ke-zheng and X.Ying, "Face Recognition Using 2DGabor Mean Values in Subspace", Journal Of Harbin University Of Science And Technology, vol. 17, no. 5, **(2012)**, pp.65-68.