A Novel Object Tracking Method based on Superpixels Cliques Appearance Model

Yi Ouyang
School of Computer and Information Engineering
Zhejiang Gongshang University
China
oyy@mail.zjgsu.edu.cn

ABSTRACT: *For robustly handling the appearance change of target object and heavy occlusion, a novel super- pixels clique based tracking algorithm is proposed. By two stage adaptive appearance modelling method, we propose the method of learning the target-background appearance framework, which is based on super pixels principle histogram bins cluster method. The process of computing superpixels cliques confidence not only store the location information of the superpixels, the super- pixels cliques recent history and last history also are equally weighted. The first phase of two-stage adaptive cliques construct and update algorithm is target template superpixels cliques construct stage. By calculating feature distance between superpixels and cliques center, it is to determine whether a superpixel belongs to the cliques. The second phase for detection and updating stage, through compare superpixels features surrounding region of target in training frame, with cliques, the confidence of cliques can be updated. For the target appearance model adaptive learning, a principle histogram bins clustering method be proposed to adaptive update appearance model, and the computational overhead is small. The object can be tracked under appearance changing and occlusion, by bayesian filtering method with using MAP and the adaptive appearance model.*

Theoretical analysis and experiments results demonstrate that our method outperforms the sate-of-the-art methods when the target under occlusion and illumination changes dramatically.

Keywords: Object Tracking, Superpixel, Classification, Feature spaces

Received: 30 October 2015, Revised 29 November 2015, Accepted 4 December 2015

1. Introduction

Target tracking is an important research area in computer vision. However, scenarios that contain the sudden appearance changes, light illumination and long lasting occlusions present such tracking with serious difficulties. Long-term real-time tracking of human action in unconstrained environments is a challenging problem. An efficient algorithm to track the target appearance model should be first considered. Based on their appearance model, the current research methods for the object tracking can be divided into two categories: generative and discriminative .

Generative classifiers learn a model of the joint probability $p(x, y)$ of the input x and the label y, and make their predictions by using Bayes rules to calculate $p(y|x)$, and then picking the most likely label y. It can use learning model to represent the main track the target, and then exploit it to search the image area, by minimizing the structural error to determine the target location. Discriminative classifiers model the posterior $p(y|x)$. Since the variations are attributed to various factors (e.g., pose change, shape deformation, illumination change, occlusion, camera viewpoint change, etc.) and cannot be seen beforehand, adopting an appropriate appearance model at tracking stage is difficult.

The most tracking methods reported to handle this difficulty, thus far, is to adaptively update the target appearance model at each frame: learn a new appearance model with time invariant characteristics extracted from historic observed target samples, and adopt the model to the current frame. Such as, IVT [1] algorithm uses a subspace model by adding adaptively modify the appearance of the model. Tracking-Learning-Detection (TLD) [2] method to track task is decomposed into three sub-processes: tracking, learning and learning, each sub- task as a separate task, each sub-task can be performed simultaneously, target tracking process along time tracking, detection operator to locate all the appearance of being tracked to determine whether the correct tracking process. Learning process is estimated detection operator error and update it in order to avoid future with the same error. Because it was not clear what kind of learning method is suitable for TLD framework will produce (1) a complex scene tracking will fail. (2) the target disappeared, detection operator continues. (3) less real-time tracking and other issues. In order to solve the above problems, Kalal [3] analyse a variety of information in video images. and proposed a new learning framework called P-N Learning for training a binary classifier from labeled and unlabeled examples. The learning process is guided by positive experts and negative experts constraints which restrict the labeling of the unlabeled set. P-N learning evaluates the classifier on the unlabeled data, identifies examples that have been classified in contradiction with structural constraints and augments the training set with the corrected samples in an iterative process.

Learning processes for any errors exist, the expert operator P-N due to mutual compensation of the error probability is limited within a certain range in order to achieve stability. Based on P-N learning method, Saigo [4] proposed PLS [4] method to obtain a better target tracking. The Fragment-based tracker [15] aims to solve partial occlusion with a representation based on histograms of local patches. The tracking task is carried out by combing votes of matching local patches using a template. There are some sparse matrix combined with particle filter for target tracking applications L1 [5].

In this paper, we use effective and efficient lower- level visual cues for object tracking with superpixels. During the training stage, the segmented superpixels are grouped for constructing a discriminative appearance model to distinguish foreground objects from cluttered backgrounds. During the test phase, a confidence map about superpixel level is computed using the appearance model to obtain the most likely target location with maximum a posteriori (MAP) estimates. The appearance mod el is constantly updated to account for variation caused by change in both the target and background. The two-stage target tracking algorithm based on principal component samples superpixels color clustering method is proposed, which is based on super- pixel target appearance mode using adaptive learning method to update. First established adaptive appearance model, the model and the target with the tracking time changes dynamically adjusted to speed up the learning speed of the model.

2. Related Work

As superpixel segmentation and object recognition can better interation, it is currently concerned by many researchers [6],[8]. The image is divided into a number of methods have significant boundary goal - Back- ground superpixels, the use of superpixel Severability be tracked. In the literature [9], we propose a tracking method based on superpixel, the tracking task into inter foreground and background segmentation, the whole process each frame independently using Delaunay triangulation decomposition, and the use of conditional random regional match, so large amount of calculation, and the method cannot handle complex scenes containing occlusion and illumination change case tracking.

Shu Wang[10] gives another super-pixel-based Bayesian tracking method SPT, the method by combining the target and background characteristics of super-pixel segmentation, and achieved more accurate tracking result- s, but due to over-superpixel feature space large, SPT method [10] used the Mean Shift clustering algorithm to cluster all the features of the training set, making the tracking efficiency is affected. LOT[11] Methods bull- dozer probability model from the EMD calculate the change with time, but algorithm fail under the target occlusion serious.

3. Superpixel Appearance Model

For superpixel tracking method, between superpixel tracking feature space larger pixel location and tracking process over issues such as the relationship is missing, the paper gives the definition of superpixel cliques, the adjacent pixel characteris tics associated with super. Firstly, through SLIC [16] method calculate the superpixels set, superpixel correlation analysis and objectives can be defined in the system with random X adjacent N and V on the grid .

Definition 1: One superpixels clique C is composed of a group of adjacent superpixels $\{X_i\}$, where $X_i \in \Omega$, and Ω is the region of superpixels $[w, h]$, $L_c(i) = (x_i, y_i)$ is the center coordinates of patch X_i .

The cliques feature can be expressed as: clique center features f_c, cliques radius R, a member of the superpixel area feature set $F = \{f_1, ..., f_N\}$ and the region of superpixels Ω four parts , $C = \{f_c, R, F, \Omega\}$. Superpixels cliques members features f is represent as color histogram , $R_c = \frac{N}{2*pi}$.

3.1 Confidence Response Map

The superpixels cliques represent the target image detection area by the joint confidence, the greater confidence degree of superpixels, the greater confidence of cliques; cliques feature super pixels farther distance from the cluster center features, the lower confidence the contributing pixels. Let w_t^c as the cliques confidence, the confidence level for the superpixel, λ as cliques 's impact factor, after the introduction of superpixel group harmony

$$\hat{w}_t = w_t^c * \lambda + w_t^i \tag{1}$$

$$w_t^i = exp(-|f_t(i) - f_c(i)|/R_c(i) \tag{2}$$

$$w_t^c = (1-\alpha)w_{t-1}^c + \alpha w_t^i \tag{3}$$

w_t^c represent the degree of in superpixels cliques interactive potential, it take account of the historic infor mation of cliques. We use exponential average method to update the weight w_t^c , and $\alpha \in [0, 1]$ as the regulatory factor. If $\alpha = 0$, and recent history has no effect (current conditions are assumed to be transient); if $\alpha = 1$, then $w_t^c = w_t^i$, and only the most recent weight matters (history is assumed to be old and irrelevant). Most commonly, $\alpha = 0.5$, so resent history and past history are equally weighted. We compare the mediate cliques confidence with superpixel confidence and cliques confidence. From the Figure 1, we can see the mediate cliques confidence not only retain the target's location characteristic, but also maintain the correlation of period.

3.2 Target Appearance Model Driven by Superpixel Cliques

Discriminative appearance model consists of two main parts: the confidence set W of superpixel cliques and the clique center features set f_c, (ie. $A_c = \{W(i), f_c(i)\}$. Appearance model using a two-stage adaptive learning approach, the specific process is as follows:

The first stage is the process of construction superpixel cliques, using SLIC [16] to perform the superpixel segmentation,which will generate $N = 300$ superpixels, and approximately normal distributed over the sample set. The color histogram features for each super pixel region was calculated, and use it to obtain the sample appearance of model $C_t(i)$, which was added to the target appearance model set $C(i)$.

Let t frame appearance model as $C_t = \{C_t(i)|i = 1, .., n\}$, we use the superpixel v_i to update C_t and generate the ith clique feature $f_c(i)$. Through superpixels cliques features distance D_s between each cliques and center clique, the cliques can be updated. If D_s smaller than threshold, then superpixels v_i should be used to update the current appearance model C_t, and use equation (1) to calculate the cliques confidence; else find the smaller distance ds from the neighbour superpixel. If the distance d_s below the clique threshold θ; then we update current clique C_t; else create the new superpixel clique C_t, the specific process as algorithm 1.

The second stage is to detect and update phase. C_{t-1} is the historical superpixel cliques set. In the t frame, we exploit SLIC to compute the superpixels set V_t, and select a superpixel v_i from V_t, computing clique confidence that the clique nearest with v_i in cliques set C_i. If the features distance between clique C_i and v_t is smaller than threshold, the clique C_i is updated by equation (3) that computing the confidence of clique; else these cliques that did not been updated can be removed.

Algorithm 1 Superpixels cliques Construction Algorithm

Require: the t frame image I_t
Ensure: denotes C_t as Superpixels set at t frame
$V_t = \{v_1, ..., v_n\}$ as the superpixels set of t frame
Using $SLIC^{[17]}$ method to calculate the image I_t superpixels set V_t
computing the color histogram features $F_t = \{f_1, f_2, ..., f_n\}$ of V_t
$V(\overline{C}) \leftarrow V_t$; select a superpixel v_i from $V(\overline{C})$;
$C_t(1) \leftarrow v_i$;$f_c(1) \leftarrow f_i$;$k \leftarrow 1$ {k as the superpixel sub};
$V(\overline{C}) \leftarrow V(\overline{C}) - v_i$;
while $V(\overline{C}) \neq \emptyset$ **do**
  random select a v_j of neigbor superpixel from $V(\overline{C})$;
  $V(\overline{C}) \leftarrow V(\overline{C}) - v_j$;
  if $v_j \neq \emptyset$ **then**
    if $\mathrm{dist}((f_c(k) - f_j) < \theta)$ **then**
      $C_t(k) \leftarrow v_j$; $f_c(k) \leftarrow \frac{1}{|C_t(k)|} \sum_{f_j \in C_t(k)} f_i$;
    else
      calculating the minimum distance $dist_j(m)$ from v_j neighbour superpixel cliques
      if $dist_j < \theta$ **then**
        $C_t(m) \leftarrow v_j$; $f_c(m) \leftarrow \frac{1}{|C_t(m|} \sum_{f_j \in C_t(m)} f_i$;
      else
        $k \leftarrow k + 1$;
        $C_t(k) \leftarrow v_j$; $f_c(k) \leftarrow \frac{1}{|C_t(k)|} \sum_{f_j \in C_t(k)} f_i$;
      end if
    end if
  end if
end while

In order to prevent degradation of cliques(the number of cliques is equal to the number of superpixels), we removed these cliques that had not been up- dated. (For woman sequence, it got 285 superpixels and had 64 cliques in #5 frame). Cliques confidence indicates that it belongs the target or background probability. Superpixel and the center of the cluster uses a weighted distance from the item to measure.

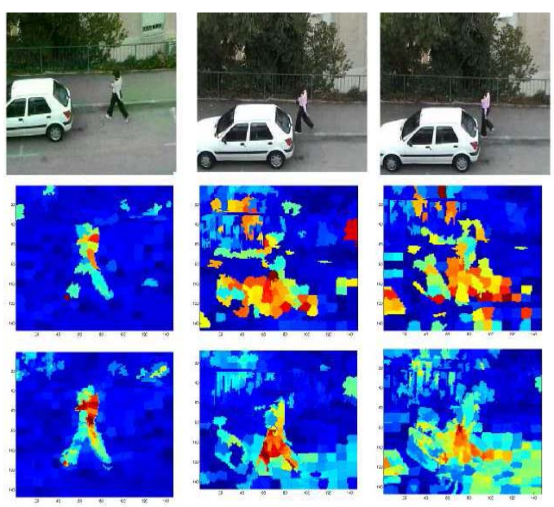

Figure 1. Confidence map. The first line are the origin images;the second line are the computed confidence map of superpixels using Eq. (2). the third line are the computed confidence map of superpixels using Eq. (3). The superpixels colored with red indicate strong likelihood of belonging to the target, and those colored with dark blue indicate strong likelihood of belonging to background

3.3 Principal Bins Features Clustering based on Superpixel Color Histogram

In order to avoid the appearance of the target and the background due to illumination, pose, occlusion and camera movements and other factors changes, this paper proposes a principal bins features clustering algorithm based on superpixel color histogram. Firstly, we select some cliques candidate set F from motion model, each clique having feature $f_i \in R^B$, where B is the number of bins(implement divide HSV into 8 bins, so $B = 512$). Firstly, we calculate the maximum bins h_{max} and the id b_{max} of the color histogram among the superpixels training set F. Let h_{max} be the principal bins component, which represent the mainly contribution in HSV color space of superpixel region. The training set is sorted in h_{max} descend order, and clustering these having same h_{max} value into one group. Finanlly, the cluster center feature can be computed from these group, and the computational overhead is $O(N + M)$, where N is the number of superpixels, M is the number of cliques. the implement process as follow:

– **Step 1:** Get maximum bin hmax and the *ID* b_{max} from each color histogram feature of superpixels in training set f_i.

– **Step 2:** Clustering maximum bin b_{max} of each super-pixels color histogram from training sets into differ- ent clusters S_k.

– **Step 3:** Computing the cluster center features

$$f_c = \sum_{i=1}^{S_n} \frac{f_i}{S_n},$$

the superpixels training set

$$S = \{\{S_k\} | f_i(b_{max}^k) \in S_k\}$$

4. Two-stage Adaptive Learning Algorithm Based On Cliques Appearance Model

The first stage is construction for the super-pixel cliques. To overcome the complexity and diversity of the scene changing, this paper use the confidence of superpixels cliques. Firstly we extract the HSV space from image color space, and use SLIC[16] method, which can produce a specified number of superpixels, to segment the image into N superpixels. According to the number of correlation superpixels and their features, the number of cliques, center feature and radius of cliques can be determined.

Algorithm 2 The Updating algorithm of Superpixels cliques

Require: the t frame imageI_t, C_{t-1} as t-1 frame superpixels cliques set;n as t-1 frame cliques number.
Ensure: C_t is superpixels cliques set
$V_t = \{v_1, ..., v_n\}$ is superpixels sets

computing histogram features $F_t = \{f_1, f_2, ..., f_n\}$ of V_t
$C_t \leftarrow C_{t-1}$
while $V_t \neq \emptyset$ **do**
  select a superpixel v_j from V_t; $V_t \leftarrow V_t - v_j$;
  $new_{clique} \leftarrow true$;
  if $v_j \neq \emptyset$ **then**
    for select a center feature $f_c(k)$ from $C_t(k) \in C_t$ **do**
      if dist$((f_c(k) - f_j) < \theta)$ **then**
        $C_t(k) \leftarrow v_j$; $f_c(k) \leftarrow \frac{1}{|C_t(k)|} \sum_{f_j \in C_t(k)} f_i$;
        $new_{clique} \leftarrow false$;
        break;
      end if
    end for
    if $new_{clique} == true$; **then**
      $n \leftarrow n + 1$;
      $C_t(n) \leftarrow v_j$; $f_c(n) \leftarrow f_j$;
    end if
  end if
end while
remove the unchanging superpixel from cliques set $C_t \leftarrow C_t - C_{t-1}^N$

Let X as random variable of a video sequence, defining the state vector $X_t = \{L_{(x,y)}, s_t\}$,where $L_{(x, y)}$ is the target center location, st is the scale factor, $p(y_t|x)$ denotes the probability of cliques belongs to target. Motion model $p\ (x_t|x_t\text{-}1)$ generate a predict x, which denotes the correlation of target's temporal structure in the video. We assume that the motion model obeys Gaussian distribution:

$$p(x_t|x_{t-1}) = N(m(x_t, x_{t-1}), \Lambda) \quad (4)$$

where Λ is the diagonal covariance matrix, $m\ (x_t|x_t\text{-}1)$ denotes the means of two random variable. We use Bayes filtering based cliques method to track target .

Prediction stage: $Bel(x) = \int p(x_t|x_{t-1})Bel(x_{t-1})dx_{t-1}$

Update stage: $Bel(x_t) \propto a_t p(y_t|x_t)Bel(x_t)$

Observation Markov assumptions: Observation depends only on the state of the current observations, i.e. $p(y_{1:t}|x_t) = p(y_t|x_t)p(z_{1:t-1}|x_t)$, the filtering process is mainly determined by the dynamic model $p(x_t|x_{t-1})$, which describes the state temporal correlation of the target between frames. Give a specific algorithm implementation process mentioned in this article as follow:

– Step 1: Construct the superpixel cliques appearance model by call the adaptive appearance model construction algorithm.

– Step 2: Segment the current frame into N superpixels, and extract their color histogram features by equation (1),(2),(3), and computing their confidence map.

– Step 3: Construct the state candidate set $\hat{X}_t$ through N times sampling.

– Step 4: Using equation (4) computing motion model $p(x_t|x_{t-1})$

– Computing target Probability $p(y_t|x_t^i) = \hat{w}_t$

– Parameter Estimation using MAP method $\hat{X}_t = argmax_{x_t^i}(Bel(x_t))$.

Sequenc	IVT	TLD	SPT	LOT	ROT	Our
Basketbal	94	7	5	6	6	5
girl mov	216	128	26	36	105	24
bolt	83	-	7	12	150	31
Liquor	54	30	8	9	34	8
Singer1	45	4	15	75	-	21
woman	161	-	11	119	113	a17

Table 1. The mean of Center D denote average errors of center position

5. Experimental Analysis

The proposed tracking method segment each video frame into different superpixels, using the HSV color space normalized histogram as the features of each superpixels, where the weighting and super pixels is 0.3 and 300 respectively. In the initial training stage, we collect training data by using the 5 frames to construct the target appearance model, and set occlusion threshold as 0.51. For cross-validation, the center position error is compared with that of current state-of-the-art methods

IVT [1], TLD [2], SPT[10], LOT [11] and ROT [12] , the executable codes of which are accessible on their own web pages. (implementation in matlab codes, In the Intel Core2 2. 4 GHz CPU). Experimental video sequences singer1 and basketball from VTD[13], lemming and liquor from PROST[14], woman from FragTrack[15] and girl mov from SPT [10].

5.1 Comparison with Several Tracking Methods

In Table-1, $D_C = \sqrt[2]{||P - P_0||}$ represents the center position coordinates errors between the tracking results and the reference standard (ground truth) , where||.|| is the Euler distance, '-' indicates partial frame detection failure in the tracking process.

Fig.3 and 5 in the target sequence occurred with the camera movement, IVT tracking will fail. Superpixel- based tracking for occlusion severe cases also make tracking results are affected (eg LOT), the key reason is no full use of the target and background appearance model, and making the tracking accuracy is limited. For considering local and environmental features surrounding , and adaptively adjusting the target appearance model in this paper, we can more accurately distinguish target and improve tracking accuracy.

Figure 2. Tracking time in a different sequence analysis

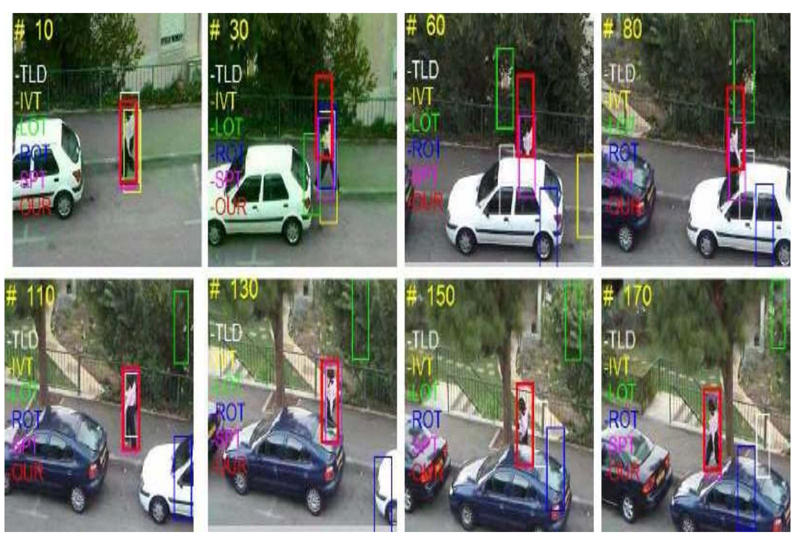

Figure 3. Tracking time in a different sequence analysis

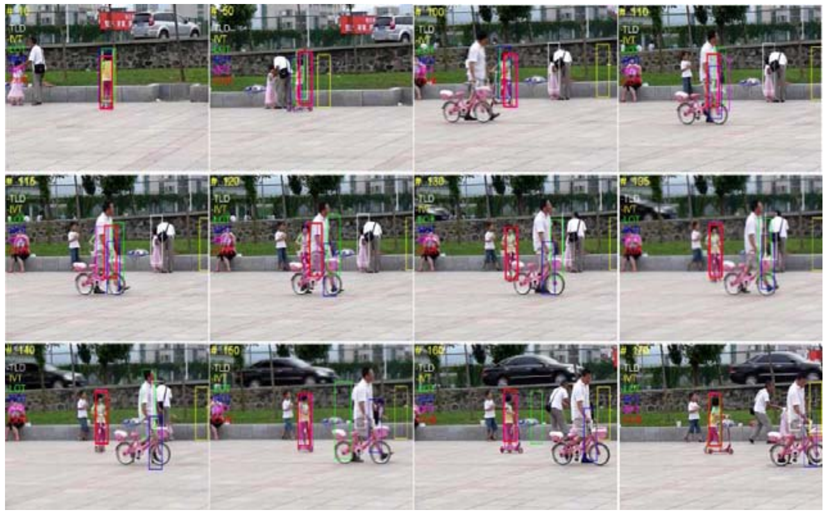

Figure 4. Tracking time in a different sequence analysis

Figure 5. Tracking time in a different sequence analysis

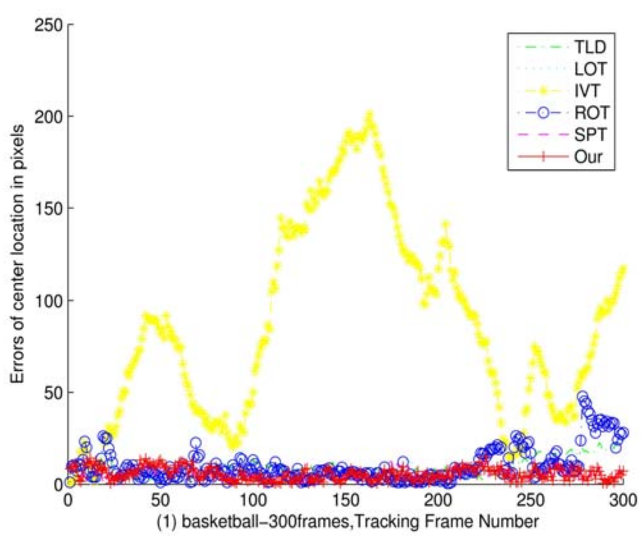

Figure 6. Tracking bounding-box center distance error in basketball sequence

5.2 Tracking under Partially Occlusion

The target objects are partially occluded in the woman (Figure 3) sequences , has a partially occlusion phenomenon. Based on SPT and the appearance of superpixel model method can eliminate the target because of changes in background appearance model for the impact, and the adaptive learning approach through dynamic update appearance models, target tracking ac-curacy of better than TLD, IVT, LOT and ROT methods to detect and deal with serious obscured targets. This test results (red) and SPT (pink) is similar. In Figure 2 basketball sequence, because the target of the block is not serious, the proposed method and SPT al-most unanimously, in Figure 4 girl mov, when traced to 115, the target as a whole is completely blocked, this method updates the target appearance model is taking a block things super pixel distribution, so the 120 is more inclined than the SPT method obstructions position, with the emergence of the target dynamically adjust the target appearance model, the tracking ac-

curacy is guaranteed. Figure 5 liquor object appears in a similar situation, the proposed method with SPT similar. The presence of the target under occlusion, the proposed method is superior to other methods with SP- T several ways. The SPT is a defect clustering model learning using MeanShift manner, so efficiency is low, the proposed method is superior in the tracking study time SPT.

Figure 7, 6, 8 show that the proposed algorithm compared with other tracking methods error distance to the center of the sequence Our algorithm (red) and SPT algorithm (pink) in the girl mov, basketball and woman sequence. The results show that it is normal condition in the center and the gound truth detection frame error is basically the same, when there is heavy occlusion, our method is significantly better than other tracking algorithms.

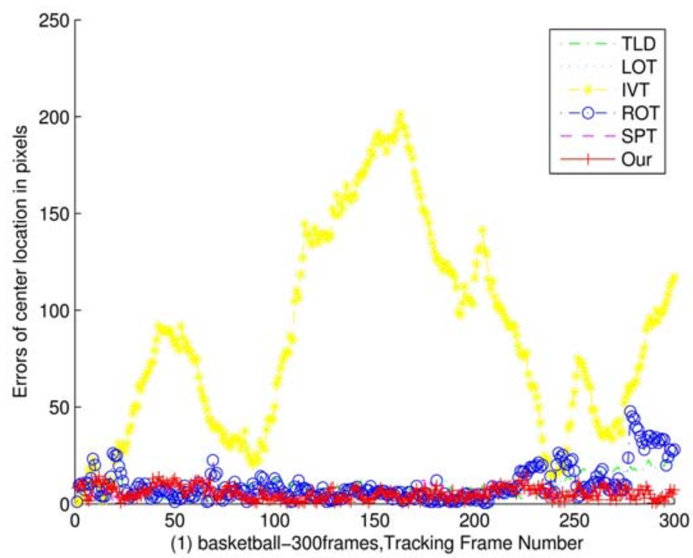

Figure 7. Tracking bounding-box center distance error in gire mov sequence

5.3 Tracking Time Analysis

As superpixel is divided with respect to the color histogram method and edge features. Our method compare with two kinds of superpixel tracking method, SPT (learning step is 5) and LOT methods. Figure 13,11,10,14 were girl mov, basketball, woman and liquor comparative sequence analysis of tracking time. Table 2 for the proposed method with SPT and LOT average tracking time per frame analysis, we can see that the proposed method is significantly faster than SPT, and better than LOT method. With the number of training samples increases, Mean Shift clustering in computation time becomes unstable, SPT method in the training time to reach the 50 seconds or more. This method of training time were less than 10 seconds. Keeping track of time, since the SPT method in the tracking stage dynamic learning updating the appearance model, so its tracking detection time was dressing and low efficiency, mainly because of the diverse set of Mean Shift Clustering on too time consuming. This adaptive learning to learn the steps of (k = 5)or below the threshold value is updated appearance model, using this method, which is much better than the SPT method, and does not lose the target and background characteristics of the distribution of super pixels.

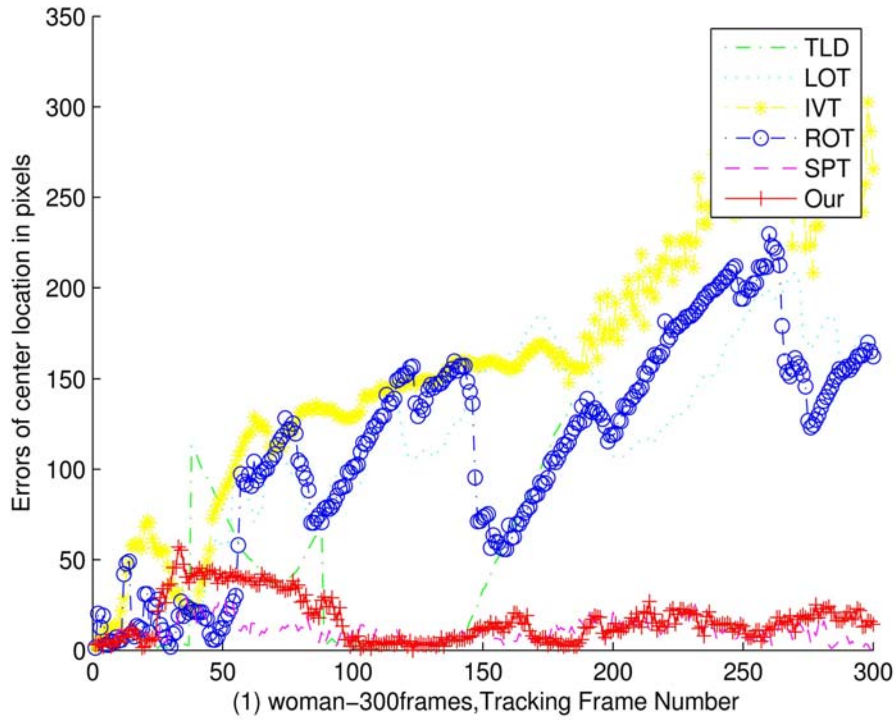

Figure 8. tracking bounding-box center distance error in basketball sequence

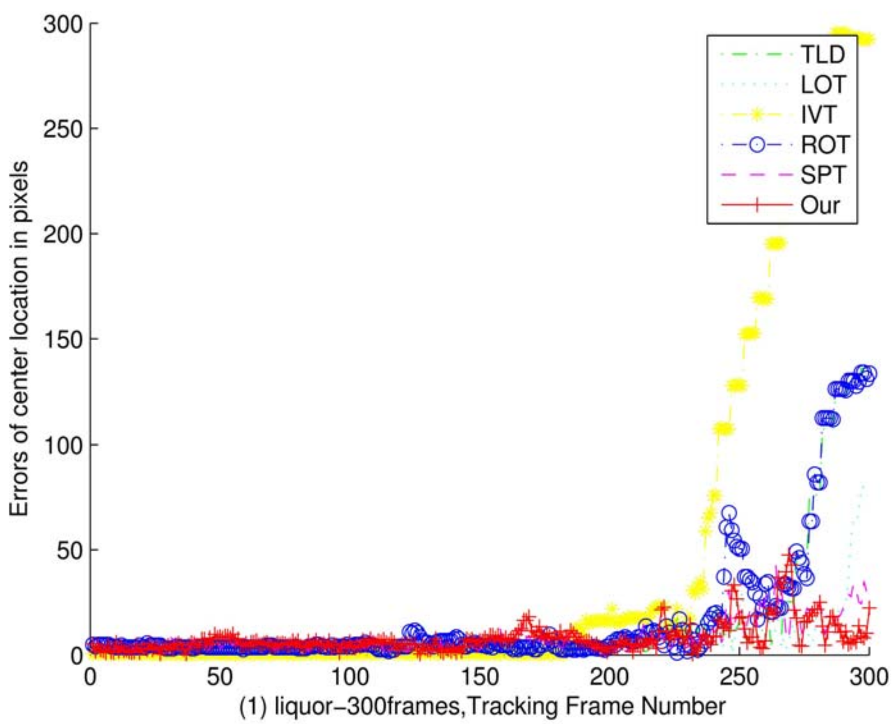

Figure 9. tracking bounding-box center distance error in basketball sequence

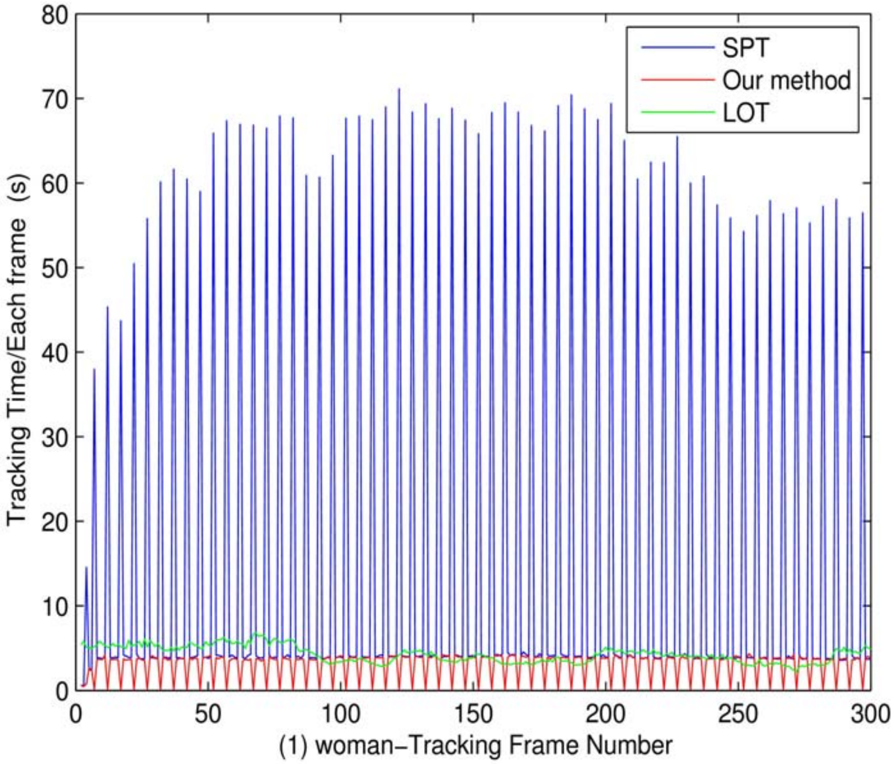

Figure 10. Tracking time in woman sequence

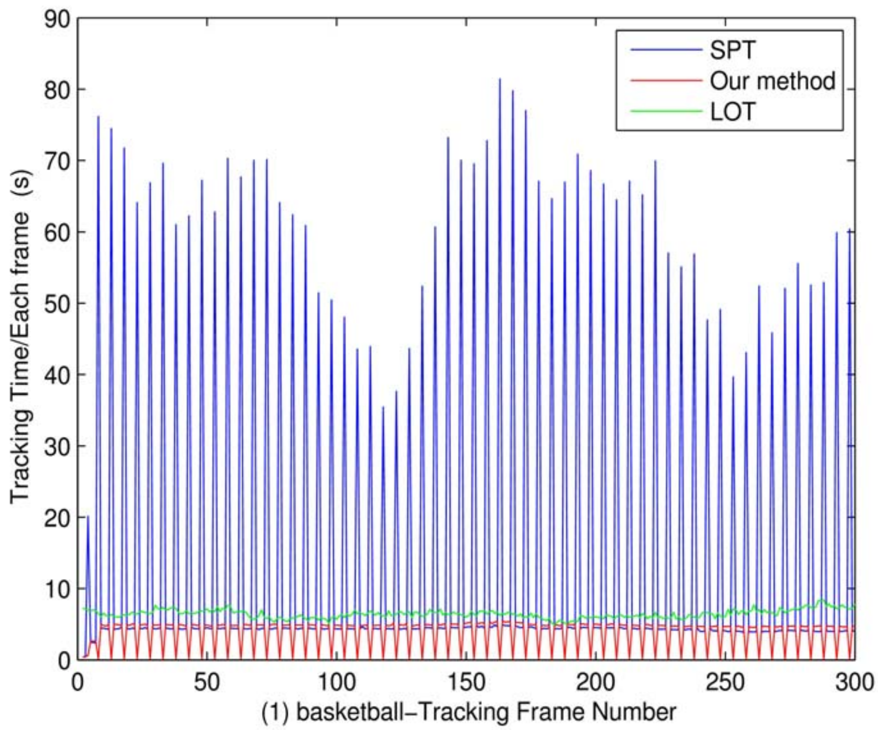

Figure 11. Tracking time in basketball sequence

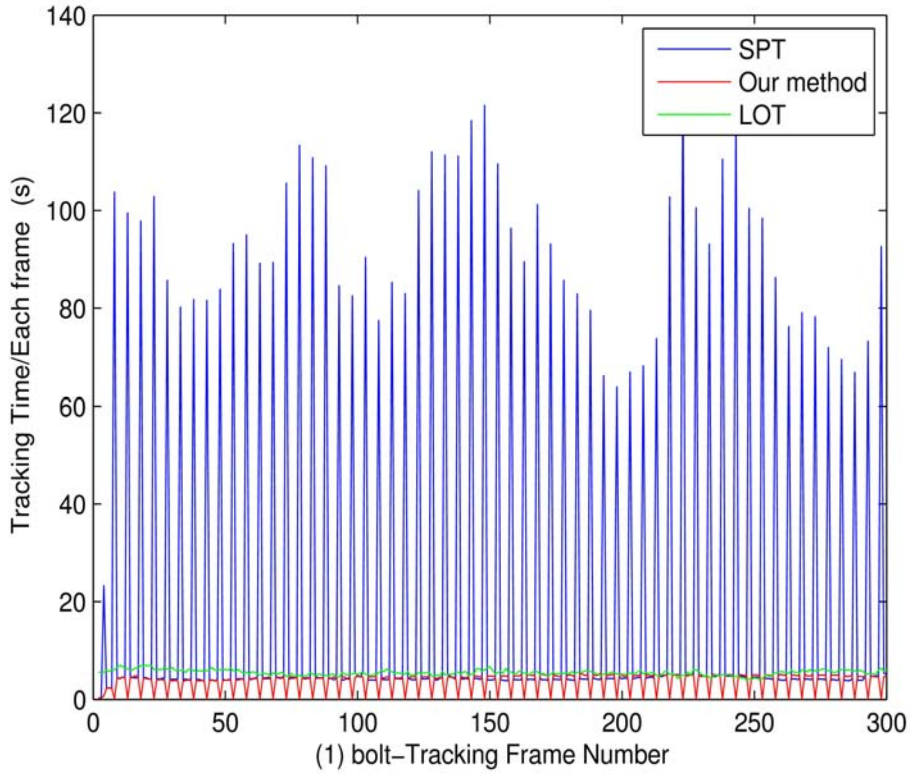

Figure 12. Tracking time in bolt sequence

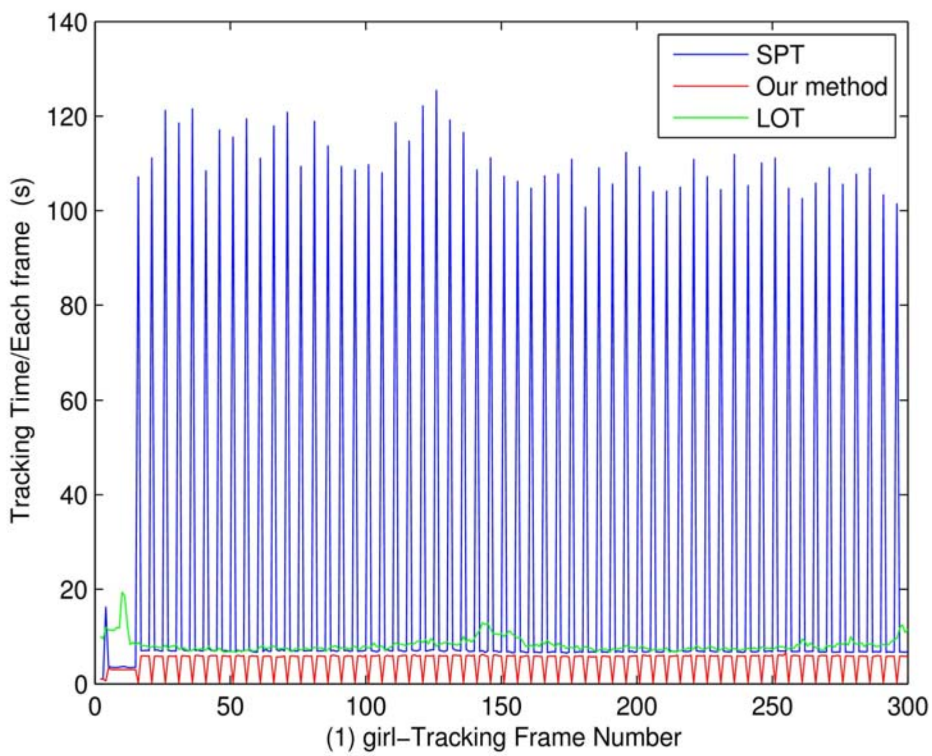

Figure 13. Tracking time in girl mov sequence

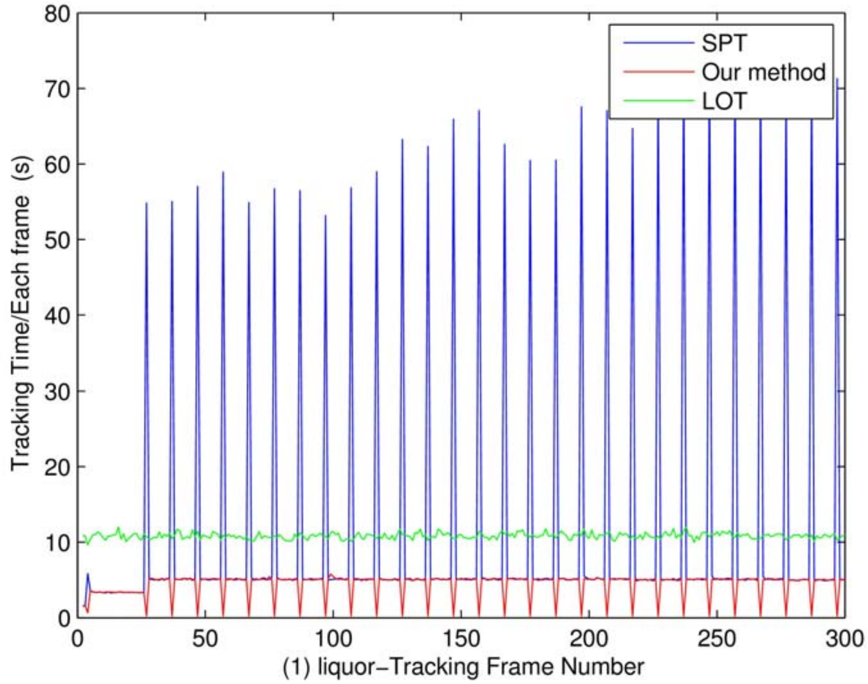

Figure 14. Tracking time in liquor sequence

SPT number	LOT number	Our method number
15.3023	4.2325	3.0631
21.5615	5.4881	3.6791
11.4861	11.095	4.5102
15.4357	6.4434	3.8534
8.7688	5.9628	4.5811
26.4612	8.225	4.6513

Table 2. Tracking methods comparison

6. Conclusion

This paper presents a novel two-sage target tracking algorithm based on superpixel cliques appearance mod- el, using the positive samples and negative samples for adaptively updated information on the target-background appearance model. To solving the these problems under abnormal situation, such as occlusion, illumination changing, and the shape changing during the process of the target tracking. We analyse the image superpixel cliques confidence response map, determine the target super-pixels in the tracking process changes in order to distinguish between target and background information. Taking into account the characteristics of the sample space is a sparse matrix, this paper uses a sample of super-pixel color main compo- nent clustering method, so always keep a sample after clustering feature in a smaller space, the use of super-pixel HSV histogram to determine the number of clusters M , removal of the smaller bins M interval, and for each super-pixel redistribu- tion class label.

Get a smaller spatial clustering, superpixel clustering algorithm time complexity is calculated, and the. We compare of this method with a variety of experiments, experiments show that our method has a strong robustness and real-time performance. The appearance model of learning time is obviously superior to SP- T method, the precision of target tracking under occlusion significantly better than the TLD, IVT, LOT and ROT methods, similar to the effect of the SPT. In multi-target tracking algo rithm on the application can be used as the next stage of the research.

References

[1] Ross, D A., Lim J., Lin, R S., et al. (2008). Incremental learning for robust visual tracking[J]. *International Journal of Comput- er Vision*, 77(1-3), 125-141.

[2] Kalal, Z., Mikolajczyk, K., Matas, J. (2012). Tracking-learning-detection, Pattern Analysis and Machine Intelligence, *IEEE Transactions on*. 34 (7), 409-1422.

[3] Kalal, Z., Matas, J., Mikolajczyk, K. (2010). Pn learning: Boot-strapping binary classifiers by structural constraints, Computer Vision and Pattern Recognition (CVPR), *IEEE Conference on*, 49-56.

[4] Saigo, H., Krmer, N., Tsuda, K. (2008). Partial least squares regression for graph mining. *In*: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 578-586.

[5] Wang Yu-Xiang., Xu Han., Ma Ting-Huai. (2012). The Aux- iliary Handoff Mechanism Based on Context and Particle Filter Algorithms[J]. *Journal of Electronics & Information Technology*, 34 (1), 223-226.

[6] Zhang, Y., Hartley, R., Mashford, J., et al. (2011). Superpixels via pseudo-boolean optimization[C].Computer Vision (ICCV), 2011 *IEEE International Conference on*, 1387-1394.

[7] Liu, M Y., Tuzel, O., Ramalingam, S., et al. (2011). Entropy rate superpixel segmentation[C]. *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on, 2097-2104.

[8] Galasso, F., Cipolla, R., Schiele, B. (2012). Video Segmentation with Superpixels[C]. *Computer Vision Asian Conference on Computer Vision*, 2013: 760-774.

[9] Ren, X., Malik, J. (2007). Tracking as repeated figure/ground segmentation[C].Computer Vision and Pattern Recognition, 2007. CVPR'07. *IEEE Conference on*, 1-8.

[10] Wang, S., Lu, H., Yang, F., et al. (2011). Superpixel tracking[C].Computer Vision (ICCV), 2011 *IEEE International Conference on*, 1323-1330.

[11] Oron, S., Bar-Hillel, A., Levi, D., et al. (2012). Locally order-less tracking[C].Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, 2012: 1940-1947.

[12] Ning, J., Zhang, L., Zhang, D., Wu, C. (2009). Robust object tracking using joint color-texture histogram[J]. International *Journal of Pattern Recognition and Artificial Intelligence*, 23 (07), 1245-1263.

[13] Kwon, J., Lee, K M. (2010). Visual tracking decomposition[C]. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 1269-1276.

[14] Santner, J., Leistner, C., Saffari, A., et al. (2010). Prost: Parallel robust online simple tracking[C].Computer Vision and Pat- tern Recognition (CVPR), 2010 IEEE Conference on, 2010, 723-730.

[15] Adam, A., Rivlin, E., Shimshoni, I. (2006). Robust Fragments based Tracking using the Integral Histogram[C].Computer Vision and Pattern Recognition 2006 IEEE Computer So-ciety Conference on, 798-805.

[16] Radhakrishna Achanta., Appu Shaji., Kevin Smith. (2010). Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, SLIC Superpixels, EPFL Technical, Report 149300, June 2010.