A Novel on-Line Tracking Method Based on Superpixels Cliques

Yi Ouyang*, Yun Ling

School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, 310018, China

* Corresponding author. Tel.: 0571-28008271; email: oyy@mail.zjgsu.edu.cn Manuscript submitted January 10, 2014; accepted March 8, 2014. doi: 10.17706/jsw.10.4.403-415

Abstract: In this paper, we propose a super-pixels clique based tracking algorithm to perform track task, meanwhile it can robustly handle the appearance change of target object and heavy occlusion. By two stage adaptive appearance modeling 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 superpixels cliques recent history and last history also are equally weighted. The first phase of two-stage adaptive cliques constructs 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. 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.

Key words: Object tracking, superpixel, classification, feature spaces.

1. Introduction

Object 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 target 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 of the input x and the label y, and make their predictions by using Bayes rules to calculate, and then picking the most likely label. 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. Since the model are influenced by various factors (e.g., pose change, shape deformation, illumination change, occlusion, camera viewpoint change, etc.), using an appropriate appearance model at tracking stage is difficult. The most tracking methods adaptively update the target appearance model at each frame, by learn a new appearance model ,which extracted from historic observed target samples. Such as, IVT [1] algorithm uses a subspace model by adding adaptively modify the appearance of the model . Tracking-Learning-Detection (TLD) [2] method

Journal of Software

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 novel 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 proposed PLS [4] method to obtain a better target tracking. The Fragment-based tracker [5] 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 [6].

In this paper, we use superpixels visual cues for object tracking. During the training stage, the segmented superpixels are grouped as cliques for constructing a discriminative appearance model to distinguish foreground objects features from cluttered backgrounds. Object tracking is posed as a binary classification problem in which the correlation of object appearance and class labels from foreground and background is modeled by partial least squares analysis, for generating a low-dimensional discriminative feature subspace purpose. As object appearance is temporally correlated and likely to repeat over time, we adaptively learn multiple appearance models for robust tracking. During the detection stage, the Superpixels Cliques(SC) confidence map is computed using the appearance model to obtain the most likelihood of target. The appearance model is constantly updated to account for variation caused by target and background changing. The target tracking algorithm based on SC model is proposed, which utilize the SC confidence dynamically adjusted to speed up the learning speed of the model during the tracking process.

2. Related Work

Object tracking has been an important topic in computer vision for several years. Recent years have been significant progress in tracking. As superpixel segmentation and object recognition can better interation, it is currently concerned by many researchers [7]-[9]. The image is divided into a number of methods have significant boundary target-background superpixels, the use of super-pixel Severability be tracked. In the literature [10], Galasso proposed a tracking method based on super-pixel, 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[11] 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 results, but due to over-superpixel feature space large, SPT method used the Mean Shift clustering algorithm to cluster all the features of the training set, making the tracking efficiency is affected. LOT[12] method use probability model from the EMD calculate the change with time, but algorithm fail under the target occlusion serious.

3. Superpixel Appearance Model

For superpixels feature space larger and tracking process templates drift these reasons, this paper gives the definition of SC at first. The adjacent target templates superpixels features construct the superpixels cliques.

Through SLIC [13] method calculate the superpixels set, the confidence map can be calculated by PLS correlation analysis.

Algorithm 1 Superpixels cliques Construction Algo-
rithm
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
1: Using $SLIC^{[17]}$ method to calculate the image I_t super-
pixels set V_t
2: computing the color histogram features F_t =
$\{f_1, f_2,, f_n\}$ of V_t
3: $V(C) \leftarrow V_t$; select a superpixel v_i from $V(C)$;
4: $C_t(\underline{1}) \leftarrow v_i; \underline{f_c}(1) \leftarrow f_i; k \leftarrow 1 \{k \text{ as the superpixel sub}\};$
5: $V(C) \leftarrow \underline{V(C)} - v_i;$
6: while $V(C) \neq \emptyset$ do
7: random select a v_j of neighbor superpixel from $V(C)$;
8: $V(\overline{C}) \leftarrow V(\overline{C}) - v_j;$
9: if $v_j \neq \emptyset$ then
10: if dist $((f_c(k) - f_j) < \theta)$ then
11: $C_t(k) \leftarrow v_j; f_c(k) \leftarrow \frac{1}{ C_t(k) } \sum_{f_j \in C_t(k)} f_i;$
12: else
13: calculating the minimum distance $dist_j(m)$ from
v_j neighbour superpixel cliques
14: if $dist_j < \theta$ then
15: $C_t(m) \leftarrow v_j; f_c(m) \leftarrow \frac{1}{ C_t(m) } \sum_{f_j \in C_t(m)} f_i;$
16: else
17: $k \leftarrow k+1;$
18: $C_t(k) \leftarrow v_j; f_c(k) \leftarrow \frac{1}{ C_t(k) } \sum_{f_i \in C_t(k)} f_i;$
19: end if
20: end if
21: end if
22: end while

Fig. 1. Superpixels cliques construction algorithm.

Definition: One superpixels clique (SC) 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 represent as : clique center features f_c , cliques radius R, a member of the superpixels 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 colour histogram, $R_c = \frac{N}{2*pi}$

3.1. Superpixels Cliques Appearance (SC) Model

Discriminative SC model consists of two main parts: the confidence set *W* and the clique center features f_c , (ie. $Ac=\{W(i), f_c(i)\}$, where *i* denotes the identify of clique). SC appearance model utilizes an adaptive learning approach. In the first stage, we use SLIC [13] to perform test frames the superpixels segmentation, which will generate *N*=300 superpixels, and approximately normal distributed over the sample set. Let *t*-*t*h frame appearance model as $SC_t=\{SC_t(i)|i=1,...,n\}$, where SC_i be generated by the *i*-*t*h clique feature $f_{c_i}(i)$. Through SC features distance *Ds* between each superpixel and center clique, the SC can be updated. If *Ds* smaller than threshold, then superpixels V_i should be used to update the current appearance model C_t ; else find the smaller distance d_s from the neighbour superpixels. If the distance d_s below the clique threshold θ , then we update current clique C_t ; else create the new superpixels clique C_t . The second stage is to detect and update phase. C_{t-1} is the historical superpixels cliques set. In the *t*-*t*h frame, we exploit SLIC to compute the superpixels set V_{t_i} and select a superpixels V_i from V_{t_i} 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 will be updated; otherwise,

these cliques will be removed.

The proposed adaptive SC appearance model is summarized in Algorithm 1. In order to prevent degradation of cliques (the number of cliques is equal to the number of superpixels), we will remove these cliques that had not been updated. (For woman sequence, it got 285 superpixels and had 64 cliques in #5 frame). SC confidence indicates that it belongs the target or background probability, we will discuss this in detail in section 3.3.

3.2. Partial Least Squares (PLS) Analysis Based on SC

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 formulate object tracking as a classification problem with PLS analysis to learn a low-dimensional and discriminative feature subspace. PLS analysis[14] is a statistical method for modeling relations between sets of variables, the observed data is assumed to be generated by a process driven by a small number of latent variables. In [15], object tracking problem is formulated as a classification with PLS analysis to lean a low dimensional and discriminative feature subspace.

Let $X \in \mathbb{R}^m$ be an m-dimensional observed variations features space and $Y \in \mathbb{R}^n$ be an n-dimensional space of labels variables. With N observed samples from each space $x \in X$ and $y \in Y$ that form two groups of variables, $X \in \mathbb{R}^{N \times m}$ and $Y \in \mathbb{R}^{N \times n}$, PLS methods find new spaces where most variations of the observed samples can be preserved, and the learned latent variables from two blocks are more correlated than those in the original spaces

$$X = TP^{T} + E$$

$$Y = UQ^{T} + F$$
(1)

where $T \in \mathbb{R}^{N \times p}$ and $U \in \mathbb{R}^{N \times p}$ are factor (score, component) matrics, $P \in \mathbb{R}^{m \times p}$ and $Q \in \mathbb{R}^{m \times p}$ are loading matrics, and $E \in \mathbb{R}^{N \times m}$ and $F \in \mathbb{R}^{N \times n}$ are residuals terms.

The PLS method, using the nonlinear iterative partial least squares(NIPALS) algorithm [14], constructs a set of weight vectors (or projection vectors) $W = \{w_1, ..., w_n\}$ such that

$$[cov(t_i, u_i)]^2 = max_{w_i=1}[cov(Xw_i, y)]^2$$

where t_i is the i-th column of matrix T, u_i the i-th column of matrix u and $cov(t_i, u_i)$ is the sample covariance between latent vectors t_i and u_i . After the extraction of the latent vectors t_i and u_i , the matrix X and vector y are deflated by subtracting their rank-one approximations based on t_i and u_i . This process is repeated until the desired number of latent vectors had been extracted. The dimension reduction is performed

by projecting the feature vector v_i , extracted from a detection window d_i , onto the weight vectors W, obtaining the latent vector Z as a result. This vector is used to classification.

Although detection results can be improved by utilizing overlapping blocks for low-level feature extraction within the detection window, the dimensionality of the feature vector becomes extremely high. As a result, the speed of the human detector decreases significantly due to the time needed to extract features and project them. To overcome this problem, we employ a two-stage approach. In a fast first stage, based on a small number of features, the majority of detection windows (those with low probability of containing humans) are discarded. The remaining windows are evaluated during a second stage where the complete set of features allows challenging samples to be correctly classified. To decompose *X* and *Y* by Equation (1.5), PLS algorithms first

(2)

compute the weight vector w_1 and c_1 such that most variations in *X* and *Y* can be retained by $t_1 = Xw_1$ and $u_1 = Yc_1$

$$w_1 = \operatorname{argmax}_{w_1} \operatorname{Var}(t_1) \tag{3}$$

where t_1, u_1 are the first columns of *T* and *U*, respectively, and Var(.) denotes the variance. Meanwhile, PLS analysis also requires t_1 to best explain u_1

$$\max_{w_1,c_1} \psi(t_1,u_1) \tag{4}$$

where $\psi(x, y) = Cov(x, y) / \sqrt{Var(x)Var(y)}$ defines the correlation coefficient between t_1 and u_1 . Therefore, w_1 and c_1 can be computed by solving the following optimization problem:

$$max(Xw_1, Yc_1)$$

s.t. $w_1^T w_1 = 1, c_1^T c_1 = 1$ (5)

The optimal weight vector w_1 for the above optimization problem is the first eigenvector of the following eigenvalue problem.

$$X^T Y Y^T X w_1 = \lambda w_1 \tag{6}$$

This first component is extracted, from which the rest of components are extracted, so that they are uncorrelated. The first component is defined as:

$$t_1 = \sum_{j=1}^p w_{ij} x_j \tag{7}$$

where x_i are the explanatory variables and y the variable to be explained.

The W_{ij} coefficients are:

$$w_{ij}^{PLS} = \frac{cov(x_j, y)}{\sqrt{\sum_{j=1}^{n} cov^2(x_j, y)}} = \frac{\langle x_j, y \rangle}{\sqrt{\sum_{j=1}^{n} \langle x_j, y \rangle}}$$
(8)

From which it can be deduced that in order to obtain W_{ij} the scalar products $\langle x_j, y \rangle$ must be calculated for each *j*. The weight vector of W reflects the importance of each original SC feature variable for object detection. If each feature variable in the selected feature space *X* is a function of superpixel location in an object region, then the importance of this feature variable is related to the discriminability between the target and the background classes at a given location. Therefore, we can use w_i^{PLS} to generate a confidence map, which shows the discriminative strength of different locations in an object region.

3.3. SC Confidence Measure



Fig. 2. Confidence map. The first line are the origin images; the second line are the computed confidence map of SC using Equation. (10). the third line are the computed confidence map of SC using Equation. (1). The red color denotes strong likelihood of belonging to the target, and dark blue indicate strong likelihood of belonging to background.

The proposed SC method is developed based on the assumption that a target image region can be better represented by the SC model. The target SC confidence response map represent the target image detection region by the joint confidence, the greater confidence value, the higher probability of target object; cliques feature super pixels far from distance from the cluster center features, the lower confidence the contributing

pixels . Let W_t^c as the cliques PLS weight, the confidence level for the super -pixel , λ as cliques 's impact factor , after the introduction of superpixels group harmony

$$w_t = w_t^c * \lambda + w_t^i \tag{9}$$

$$w_t^i = exp(-|f_t(i) - f_c(i)| / R_c(i))$$
(10)

$$w_t^c = (1 - \alpha) w_{t-1}^c + \alpha w_t^i$$
(11)

 w_t^c represent the degree of in SC potential, it take account of the spatial information of cliques. We use exponential average method to update the weight w_t^c , and $\alpha \in [0,1]$ as the regulatory factor. $\alpha = 0$ denotes 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, $a = 0.5^{'}$, so resent history and past history are equally weighted. We compare the mediate cliques confidence with SC confidence and cliques confidence. In the Fig. 2, the mediate cliques confidence not only retain the target's

location characteristic, but also maintain the correlation of period.

4. Adaptive Learning Algorithm Based on SC Model

Since the appearance change of an object during a long period of time may be quite nonlinear and complex, one linear appearance model is not likely to suffice. However, appearance of a target object may be temporally correlated and may repeat over time. Inspired by [16], we therefore learn multiple appearance models for more effective object representation. Fig 2 shows that observations of a target over a long period can be divided into multiple sets. Within the *i-th* set, the object appearance dosed not change much and we use superpixels cliques to

learn a discriminative appearance model $A_i = \{x_i^p, x_i, W_i\}$. Therefore, the appearance of a target object can be

represented by multiple appearance models $A = \{A_1, ..., A_k\}$, where k is the number of appearance models. The proposed representation scheme is more effective than existing methods base on single linear appearance

model. Note that only one appearance model A_1 is initialized in the first frame. With our representation scheme, we define a distance metric for the tracking task. In this paper, the distance between a test sample

 $x \in \mathbb{R}^m$ and the learned appearance model set A is defined as

$$Dist = \sum_{i=1}^{k} ||W_{i}^{T}(x - \overline{x}) - W_{i}^{T}(x^{p} - \overline{x})||_{2}^{2}$$
(12)

where x^p is the mean of the positive samples used in training $A_i = \overline{x_i}$ is the mean of all the samples in training A_i , and $\|.\|_2$ is Euclidean norm. The target and background appearances may change due to factors such as illumination, pose, occlusion, camera motion, and so on. To deal with this problem, we propose an set adaptive object representation method. Let the current of appearance models be $A = \{A_i \mid i = 1, ..., k, k \le K\}$. When the tracking result at time t is obtained, we use the corresponding target observation x_t to update A. Since we have computed the distances from the target observation x_t to all the appearance models in A for determining the tracking result, we select the appearance model $A_{\rm e}$ with the smallest distance distance d_s and the appearance model A_l with the largest distance \$d'\$. If d_s is less than a predefined threshold, x_t is utilized to update A_s .

The update process includes three components: the mean of the positive sample $\overline{x_i^p}$, the mean of all the training samples \overline{x}_i and the weight matrix $W_{\perp} \overline{x^p}$ can be updated by using a a random update probability. Both \hat{x} and W can be updated by SC Construction method with the positive and negative samples. If $||A_s - A_t||$ is larger than the predefined threshold and k < K, a new appearance model A_{k+1} is added to A. If $||A_s - A_t||$ is larger than the predefined threshold and k = K, a new appearance model is initialized to replace A_t in A.

5. Object Tracking by SC Model

To overcome the complexity and diversity of the scene changing, this paper uses the confidence of superpixels cliques. Firstly we extract the HSV space from image color space, and use SLIC[13] 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.

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, S_t is the scale factor, $p(y_t | x)$ denotes the probability of cliques belongs to target. We note the tracking result is the MAP estimation over the samples ,which can be modelled well with mode

seeking. Motion model $p(x_t | x_{t-1})$ generate a predict x_t^i , which denotes the correlation of target's temporal structure in the video. We assume that the motion model obeys Gaussian distribution :

$$p(x_{t}^{i} | x_{t-1}) = w_{t}^{i} \times G(x_{t-1}, \Lambda)$$
(13)

where Λ is the diagonal covariance matrix, x_t^i denotes the weight of the *i*-th sample at frame computed by the corresponding template set. We use Bayes filtering based cliques method to track target.

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

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

 $p(x_t \mid x_{t-1})$, which describes the state temporal correlation of the target between frames .

Give a specific algorithm implementation process mentioned in this paper 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 X_{t} through *N* times sampling.

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

Computing target Probability $p(y_t | x_t^i) = w_t$

Parameter Estimation using MAP method $X_t = argmax_{x_i}(Bel(x_t))$



Fig. 3. Tracking results in a basketball sequence analysis.

Journal of Software



Fig. 4. Tracking results in a woman sequence analysis.



Fig. 5. Tracking results in a girl_mov sequence analysis.



Fig. 6. Tracking results in a liquor sequence analysis.

6. 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 SC 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[11], LOT[12] and ROT [18], the executable codes of which are accessible on their own web pages. (implementation in the Intel Core2 2. 4 GHz CPU). Experimental video sequences singer1 and basketball from VTD[19], lemming and liquor from ROT[18],

woman from FragTrack [5] and girl_mov from SPT [11].

Table 1. The mean of Center D Denote Average Errors of Center Position IVT TLD SPT LOT ROT Sequenc Our **Basketball** 94 7 5 6 6 5 girl mov 216 128 26 36 105 bolt 83 7 12 150 31 Liquor 54 30 8 9 34 8 Singer1 45 4 15 75 21 Woman 161 11 119 113 17

6.1. Comparison with Several Tracking Methods

In Table 1, $D_c = \sqrt[3]{|P-P_0||}$ 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. 4 and Fig. 6 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.



Fig. 7. Tracking bounding-box center distance error in basketball sequence.

6.2. Tracking under Partially Occlusion

The target objects are partially occluded in the woman (Fig. 4) sequences , has a partially occlusion phenomenon. Based on SPT and the appearance of super-pixel 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 accuracy 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 Fig. 2 basketball sequence, because the target of the block is not serious, the proposed method and SPT almost unanimously. In Fig. 5, 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 accuracy is guaranteed. Fig. 5 liquor object appears in a similar situation, the proposed methods with SPT several ways. The SPT is a defect clustering model learning using Mean Shift manner, so efficiency is low, the proposed method is superior in the tracking study time SPT. Fig. 7- Fig. 9 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.



Fig. 8. Tracking bounding-box center distance error in gir_mov sequence.



Fig. 9. Tracking bounding-box center distance error in woman sequence.



Fig. 10. Tracking bounding-box center distance error in liquor sequence.

7. Conclusion

This paper presents a novel robustly tracking algorithm based on SCA model, 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 analyze the image superpixels cliques confidence response map, determine the target super-pixels in the tracking process changes in order to distinguish between target and background information. This paper propose a superpixels clique approach to improve the

robustness and accuracy of object tracking. For tackle occlusion and illumination problems, this paper presents a novel two stage adaptive appearance modeling method. The confidence map of superpixels cliques not only store the spatial information of the superpixels, the temporal information also are weighted. To overcome the template drift, this paper proposes partial least squares(PLS) analysis method to generate a low-dimensional discriminative feature subspace with SC appearance model for robust tracking. The PLS weight vector reflects the importance of each SC feature variable for object detection.. The appearance model of learning time is obviously superior to SPT 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 algorithm on the application can be used as the next stage of the research.

Acknowledgment

The authors would like to thanks the anonymous reviewers for their constructive and useful comments, which helped in improving the presentation of our work. This work was supported by the Ministry of Education of China under Grant No. 13YJCB60043.

References

- [1] Ross, D. A., Lim, J., Lin, R. S. *et al.* Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3).
- [2] Kalal, Z., Mikolajczyk, K., & Matas, J. (2012). Tracking-learning-detection, pattern analysis and machine intelligence. *34(7)*, 1409-1422
- [3] Kalal, Z., Matas, J., & Mikolajczyk, K. (2010 Bootstrapping binary classifiers by structural constraints. *Proceedings of the 2010 IEEE Conference on Computer vision and pattern recognition* (pp. 49-56).
- [4] Saigo, H., Krämer, N., & Tsuda, K. (2008 Partial least squares regression for graph mining. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 578-586).
- [5] Adam, A., Rivlin, E., & Shimshoni, I. (2006 Robust fragments-based tracking using the integral histogram. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 798-805).
- [6] Wang, Y. X., Xu, H., & Ma, T. H. (2012). The auxiliary handoff mechanism based on context and particle filter algorithms. *Journal of Electronics and Information Technology*, *34*(1), 223-226.
- [7] Zhang, Y., Hartley, R., Mashford. J. *et al.* (2011). Superpixels via pseudo-boolean optimization. *Proceedings of the 2011 IEEE International Conference on Computer Vision* (pp. 1387-1394).
- [8] Ren, X., & Malik, J. (2007). Tracking as repeated figure/ground segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8).
- [9] Liu, M. Y., Tuzel, O., Ramalingam, S. *et al.* (2011). Entropy rate superpixels segmentation. *Proceedings of the* 2011 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2097-2104).
- [10] Galasso, F., Cipolla, R., & Schiele, B. (2012). Video Segmentation with Superpixels. *Proceedings of the* Asian Conference on Computer Vision 2012 *Computer Vision* (pp. 760-774).
- [11] Wang, S., Lu, H., Yang, F. et al. (2011). Superpixels tracking. Proceedings of the 2011 IEEE International Conference on Computer Vision (pp. 1323-1330).
- [12] Oron, S. et al. (2012). Locally orderless tracking. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition(pp. 1940-1947).
- [13] Radhakrishna, A., Appu, S., Kevin, S., Aurelien, L., Pascal, F., & Sabine, S. (2010). *SLIC Superpixels*, EPFL Technical Report.
- [14] Wold, H. (1985). Partial least squares. *Encyclopedia of Statistical Science*, New York: Wiley.
- [15] Wang, Q., Chen, F., Xu, W. *et al.* Object tracking via partial least squares analysis, *Image Processing*, *21(10)*, 4454-4465..
- [16] Ning. J., Zhang, L., Zhang, D., & Wu, C. (2009). Robust object tracking using joint color-texture histogram.

International Journal of Pattern Recognition and Artificial Intelligence, 23(7). 1245-1263.

- [17] Benito, C. C., & Dan, S. (2014). Parameter estimation using polynomial chaos and maximum likelihood. *International Journal of Computer Mathematics*, *91(2)*, 336-346.
- [18] Santner, J., Leistner, C., Saffari, A. *et al.* (2010). Prost: Parallel robust online simple tracking. *Proceedings of the* 2010 IEEE Conference on Computer Vision and Pattern Recognition (pp. 723-730).
- [19] Kwon, J., & Lee, K. M. (2010). Visual tracking decomposition. *Proceedings of the* 2010 IEEE Conference on *Computer Vision and Pattern Recognition* (pp. 1269-1276).



Yi Ouyang's major research areas focus on visual computing subjects in the areas of computer vision and pattern recognition. He has researched on object detection, tracking, pose estimation, and recognition. Potential applications of his research reside in the domains of human-computer interaction, visual surveillance, and video analysis and content extraction.



Yun Ling is a full professor. He focuses on visual tracking and segmentation. His current research interests include computer vision and pattern recognition. He has led many national and international research projects and industrial projects. He has researched on the ways of analyzing the performance of computing grid and put forward a range-based approach to distributed computing performance.