www.s2is.org

International Journal on Smart Sensing and Intelligent Systems

ISSN 1178-5608

ROBUST VISUAL TRACKING BASED ON SUPPORT VECTOR MACHINE AND WEIGHTED SAMPLING METHOD

Gao Xiaoxing and Liu Feng

Shijiazhuang University of Economics, Shijiazhuang 050031, China

Submitted: Oct. 5, 2014 ***Accepted: Jan. 12, 2015*** ***Published: Mar. 1, 2015***

Abstract- Visual tracking algorithm based on binary classification has become the research hot issue. The tracking algorithm firstly constructs a binary classifier between object and background, then to determine the object's location by the probability of the classifier. However, such binary classification may not fully handle the outliers, which may cause drifting. To improve the robustness of these tracking methods, a novel object tracking algorithm is proposed based on support vector machine (SVM) and weighted multi-sample sampling method. Our method constructs a classifier by sampling positive and negative samples and then to find the best candidate that has the largest response using SVM classifier. What's more, the proposed method integrates weighted multi-instance sampling method, which can consider the sample importance by the different weights. The experimental results on many sequences show the robustness and accuracy of the improved method. The proposed target tracking algorithm in video target tracking with a variety of classic popular tracking algorithm, better able to achieve robust target tracking, but also in the infrared video, the infrared target tracking is also has the advantages of stable and accurate..

Index terms: Visual tracking; support vector machine (SVM); weighted multi-sample sampling method.

I. INTRODUCTION

Visual object tracking is a comprehensive technique for locating and tracking object by analyzing and understanding video information captured by visual sensor. It has an important status in the field of computer vision. In the research and application relating to visual object tracking, robustness is the most basic and important problem. With rapid development of computer hardware technique, real-time ability of visual object tracking algorithms rely more and more on hardware mechanism. The pivotal issue of visual object tracking is to improve the robustness of dealing with all kinds of disturbances in all kinds of environments. Visual tracking has applied in intelligent video surveillance, intelligent traffic regulation, medical diagnostics, video-based human-computer interaction and so on [1-5].

Intelligent video surveillance including banks, shops and supermarkets, car parks, government offices, kindergartens and schools, and other places battlefield environment monitoring. U.S. 911, London subway bombings, the Russian human bombings and other terrorist incidents of intimidation send a signal to the public safety, and constantly threaten public safety. States have put security monitoring issues on the agenda, put a lot of research funding monitoring method, to develop monitoring equipment. Among them, the intelligent video surveillance with its higher cost, concealment and anti-electronic jamming ability of the advantages of being more concerned about. Traditional video surveillance need to monitor staff time watching video scenes and understand events, however, with the increase in monitoring sites, monitoring expand the scope of this approach is difficult to meet the real-time monitoring requirements. The intelligent video surveillance in real time video content analysis and processing, automatic target detection, identification, classification and tracking, to provide information for senior video analysis and understanding, to achieve automatic detection of suspicious targets and suspicious incidents automatic alarm. Among them, the real-time target tracking and robust technology is one of the key technologies of intelligent video surveillance.

Intelligent traffic regulation: With the growing number of countries in the vehicle, worsening congestion, people travel severely affected, a lot of valuable time wasted, for which many countries have built a set of video-based vehicle and pedestrian detection, intelligent road traffic grooming, suspicious vehicle tracking and positioned in one of the intelligent video regulatory network, which has important practical significance and economic value to prevent accidents, reduce traffic congestion, real-time event processing.

Medical diagnostics: To promote the development of medicine and improve medical standards, improve medical conditions, reduce medical harm, equipped with a variety of video sensors miniature medical equipment continuously available. Miniature robot enters the body lesion site, lesion analysis by video technology traits, always locate the lesion area, assisting doctors or complete autonomy precise surgical tasks, thereby effectively improving the accuracy of clinical surgery, which greatly reduce the patient's body injury. Long-term follow-up observation by video microscopic cell activity status, analyze its traits and characteristics, provide an important basis for its pathological studies.

Video-based human-computer interaction: In recent years, with the development of virtual reality technology, computer keyboard and mouse-based human-computer interaction is no longer satisfy people's ease of operation and communication intelligence requirements, more natural, more humane, more freedom of human-computer interaction gradually become the dominant way to achieve this goal, the key to let the computer can detect body posture, analyze and understand people's intentions. Video-based human-computer interaction through video content analysis and processing, the computer not only see people's behavior and understanding people's intentions. From a technical point of view, of human motion tracking is one of the key technologies to achieve this goal.

Currently, the classification has been successfully applied to track the target. Visual tracking algorithm based on binary classification has become the research hot issue. Grabner et al. designs an efficient online learning Boosting classifier [6, 7], its main principle is based on the adaptation to choose different characteristics, target effectively separated from the background, and the classification is through its continuous learning to real-time updates. When the target changes in appearance were strong, the classifier still better classification characteristics. Then a semi-supervised target tracking method is proposed [8-10], they put the tracking problem as a semi-supervised learning problem, but the disadvantage of this method are: the situation appears to block the drift phenomenon, and for semi-supervised learning priori information may only use some features specific goals and objectives. Stalder et al [11-13] also proposed a semi-supervised learning better tracking method to solve the drift problem. To resolve the ambiguity sampling samples, Babenko et al. [14-16] proposed a multi-instance learning target tracking algorithm, the method of the sampling location near the target sample as a positive sample bag, and away from the target position of the sample, respectively. Tracking is as a binary classification problem, the method achieved better tracking performance [17-19].

The closer the distance between the current position of the target's center is more likely to collect samples of the next frame is the target position, the current position that is farther from the sample to the low contribution to the classifier, and the closer the target sample the current location of the greater contribution to the classifier [20-23]. Therefore, this is now an improved weighted multi-instance learning sample sampling method. This method is added to the positive samples collected from the corresponding weights according to their distance from the center position, so that each sample has different contribution to classifier according to its weight. In this paper, we propose a novel tracking method based on SVM and weighted multi-sample sampling method. Our method constructs a classifier by sampling positive and negative samples and then to find the best candidate that has the largest response using SVM classifier. What's more, the proposed method integrates weighted multi-instance sampling method, which can consider the sample importance by the different weights. Experimental results on five challenge sequences show the accuracy and robustness of our proposed method.

This rest paper is organized as follows. Section 2 concisely introduces the basic algorithm of support vector machine. Then we introduce the proposed method with weighted multi-sample sampling method in details in Section 3. Experimental results are drawn in Section 4 and conclusions are described in Section 5.

II. RELATED THEORY OF SUPPORT VECTOR MACHINE

2.1 Support vector machines theory

Support vector machines (Support Vector Machine, SVM) is proposed [14, 15] in 1995 for the first time. SVM generalization capability, especially for small sample size problem, multi-linear, non-linear and high-dimensional pattern recognition problem, it can show a unique advantage. The idea of using SVM classification are: linear inseparable for sample collection, you need to first sample through a nonlinear transformation is mapped to high dimensional feature space, finding an optimal hyper plane in high dimensional feature space, to obtain the best classification performance. Suppose there is a class of sample data sets:

$$(x_1, y_1), (x_2, y_2), \cdots \qquad y_i \in \; - \tag{1}$$

To implement the nonlinear transformation $Z =$ for the sample data x_l

$$\begin{cases} + \geq \quad = \\ \leq - \quad = - \end{cases} \tag{2}$$

where w represents the weight vector and b is offset vector.

Transform the above equation to obtain:

$$y_i(w^T z_i + \quad \geq \quad = \quad \cdots \tag{3}$$

Assume the optimal classification plane equation is expressed as

$$w_0^T z + \quad = \tag{4}$$

Classifier interval

$$\rho \quad = \quad _{\{x|y=1\}} \frac{\ ^{T}}{\|w\|} - \ _{\{x|y=-1\}} \frac{\ ^{T}}{\|w\|} \tag{5}$$

w_0 must meet

$$\rho \quad = \frac{}{\|w_0\|} = \frac{}{\|w_0' w_0\|} \tag{6}$$

To obtain the maximum, the general quadratic programming is used to solve the optimization problem:

$$\max_{w,b} \phi \quad = \frac{1}{2} \tag{7}$$

The constraints of the above formula is $y_i(w^T z_i + \quad \geq \quad = \quad \cdots$.

To solve this problem linearly inseparable, we first construct a Lagrangian constraints. Therefore, the optimization problem is to solve the minimum of the formula:

$$\phi \quad = \frac{1}{2} \quad + \sum \tag{8}$$

Where ς classifier error. Then to obtain:

$$w_0 = \sum \tag{9}$$

The decision function is written as:

$$f = \quad \sum \quad + \tag{10}$$

To deal with nonlinear SVM classification problem, we need to study the problem of low-dimensional feature space into a high dimensional feature space, so that the data can be divided into a linear problem in high-dimensional space. Role of nuclear function is to map data from low-dimensional space to a high-dimensional space, without calculating the mapping function, only concerned with the output value of its classification after. Therefore, the use of kernel function method to rewrite the formula in the decision function [16-18, 25]:

$$f = \sum \quad + \tag{11}$$

2.2 Tracking algorithm for support vector machine based on vision

The premise can be divided into feature points in a linear, we had these high-dimensional space of points to do linear classification, then can find separately two hyperplane, and support vector from the nearest, and the maximum distance between the two plane. As shown in Figure 1, the 2/|w| is easy to get between the two distance to the hyperplane, we get the maximum interval, namely the need to minimize |w|, at the same time, our team of sample points add about |w| beam conditions, so that the sample data points outside the spacer hyperplane.

$$w\cdot \quad - \quad \geq \tag{12}$$

$$w\cdot \quad - \quad \leq - \tag{13}$$

The two equations can be written as:

$$c_i(w\cdot \quad - \quad \geq \quad \leq \; \leq \tag{14}$$

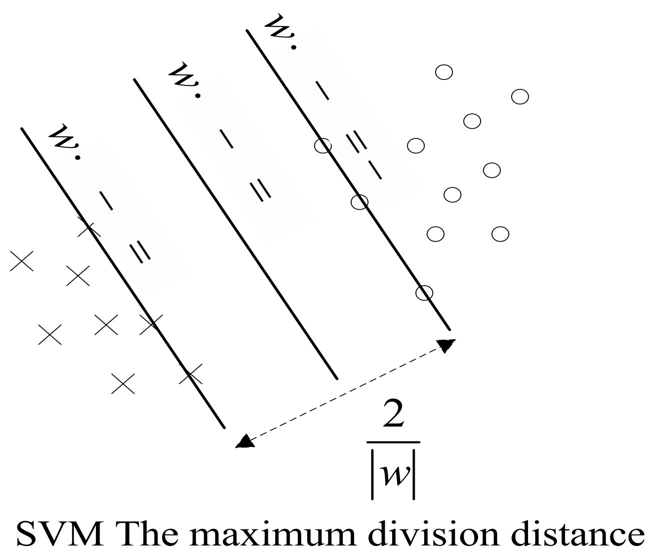


Figure 1. The maximum division distance of SVM

Support vector machine classification of samples is by finding a maximum interval of sample (Figure 1) can be distinguished hyperplane to complete, it will be the first samples are mapped to high dimensional space, and then find the maximum interval between samples of the hyperplane, and classifier error and the maximum spacing of ultra flat surface is proportional to the distance.

III. OUR PROPOSED OBJECT TRACKING METHOD

3.1 Weighted multi-sample sampling method

In the classical binary classification problem based target tracking algorithm, the first feature by sampling the target as a positive sample, the sampling background information as a negative sample, and then training a classifier to obtain positive and negative samples, sample sampling multiple locations in the next frame, then the position of the target obtained by the best classifiers. Sampling method is usually the target sample is extremely small neighborhood of the current position of the sample as a positive sample, the sample away from the current target position as the negative samples. The sampling method as shown in Fig.2 (a), to assume that l_t represents the location of the target at t-th frame, we mark $l(x)$ represents the location of each image sub-patch in the neighborhood of the current position. Therefore, sample positive samples meets the following criterion:

$$X^p = \qquad - \qquad < \tag{15}$$

where X^p is the set of samples set, s represents the search radius that is relative to the current target position. To sample negative samples meets the following criterion:

$$X^n = \qquad - \qquad < \tag{16}$$

However, the above-described analysis method has an obvious defect in the acquisition of a positive sample is collected in a sliding window of the current target position of a small neighborhood, according to the motion characteristics of the target, the target position from the actual sample increasingly unlikely that the most the next frame is called a target position, that is, the closer the distance between the current position of the target's center is more likely to collect samples of the next frame is the target position, the current position that is farther from the sample to the low contribution to the classifier, and the closer the target sample the current location of the greater contribution to the classifier. Therefore, this is now an improved weighted multi-instance learning sample sampling method. This method is added to the positive samples collected from the corresponding weights according to their distance from the center position, so that each sample has different contribution to classifier according to its weight.

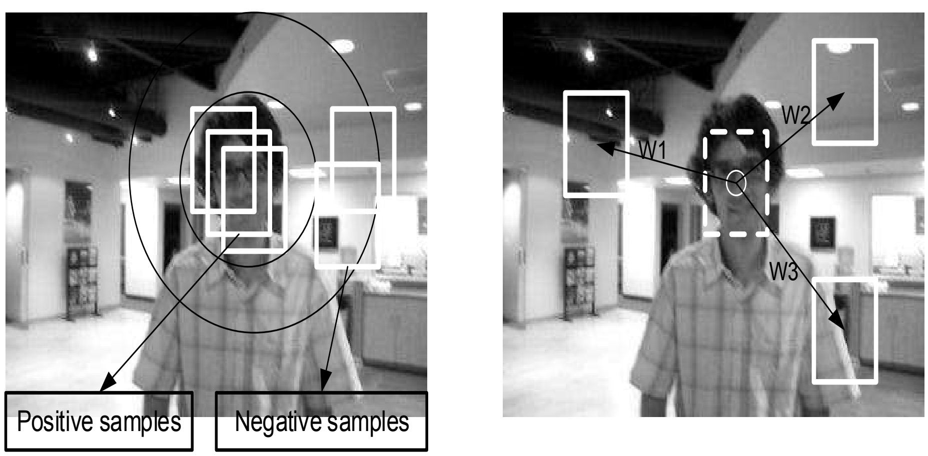
Fig.2 (b) shows the proposed weighted multi-sampling method, in which the dashed box is the current location of the target is located, for the rest of the three sampling window size according to their distance from the dashed box corresponding weights given value. Therefore, the weights is calculated as:

$$w_i = \frac{1}{\|l_t - \quad} \tag{17}$$

To normalize:

$$w_i = \frac{\quad}{\sum} \tag{18}$$

where $l(x_i)$ is the location of the i-th sample, and N is the number of samples.



(a) original sampling method (b) weighted sampling method

Figure 2. Comparison of different sampling methods

3.2 The proposed tracking method

Harr features is used to extract object features. The method of the sample is calculated for each image block obtained Harr feature vector collection, each sample is represented by a collection of a column vector. The main steps in the proposed tracking algorithm are followings. Firstly, we obtain the original location in the first frame by manually or detect the true target location, then sampled a number of positive and negative samples in a small neighborhood of the current position, and calculate the weight of each sample, and then enter a SVM classifier for the next frame, according to the random drift model and the particle filter thinking, in the vicinity of the target location on a random sampling frame multi-stem template samples and calculate the corresponding weights. Finally, the biggest response classifier will be as the optimum target position of the next frame. After obtaining the optimal position needs to be updated online classifier, the classifier based on template sample response, discarding the smaller sample value of the response, while retaining the larger sample response training continues to generate a new classifier, repeating the above processing.

For each object in the T-1 time, the deterministic in some sports constraints define it with the t moment an object corresponding to the cost function. Tracking is designed in the corresponding relationship between all (Figure 3a) to get a one-to-one correspondence between the (Figure 3b), there are usually makes the corresponding value of the cost function method for solving minimum Hungary algorithm and greedy search algorithm.

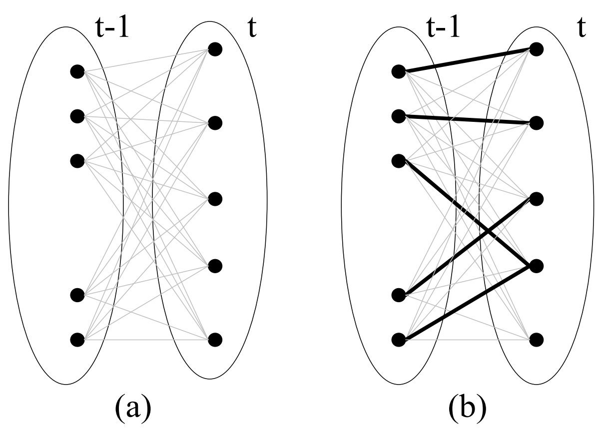


Figure 3. Tracking designed in the corresponding relationship

1. Approximate location: not hypothetical objects between frames change greatly

2. The maximum speed: defines the maximum value of the velocity of an object thus limiting the scope of the object corresponding to the

3. Smooth motion direction and velocity of movement of the object: the hypothesis does not change significantly

4. Public sport: a small region point motion is similar, this assumption is suitable for objects with multiple points representation of the situation

5. Rigid: the assumption that the objects are rigid, i.e. the distance between any two points on the object is fixed

3.3 Propagation of discrete time state probability

The propagation probability density is divided into three steps, respectively is the offset. Diffusion and estimation was shown in Figure 4. Component to determine the probability density in the whole probability distribution in a certain direction, unified offset, and the random component to its probability distribution brought some uncertainty, which we call the diffusion. Finally, some observation model help estimate of probability distribution.

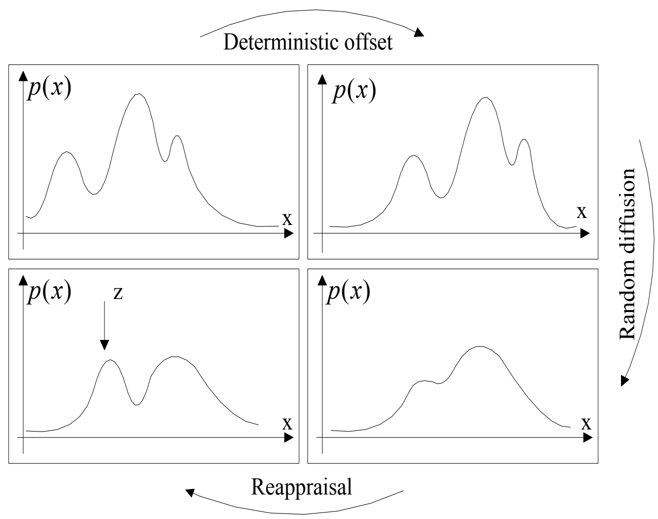


Figure 4. Diffusion and estimation

In order to calculate the need, transmission probability density must be performed in the discrete time t. Let t moment to express state variables of the target object is St, the sequence is $X_t = \ldots, S_t$ value of its history, to observe the value of the state variables of it, its historical value sequence is $O_t = \ldots, I_t$, note that here we do not have on the probability distribution of the state variables make any assumptions (such as linear, Gauss distribution).

IV. EXPERIMENTS AND ANALYSIS

We evaluate the proposed tracking method based on SVM and weighted multi-sample sampling method using five video sequences with impacted factors including abrupt motion, cluttered background, severe occlusion and appearance change. We compare our proposed tracker with other two tracking methods including Mean-shift method and original tracking method based on SVM. Fig.5, Fig.6, Fig.7, Fig.8 and Fig.9 show the tracking results on the different sequences, where each row is different from the tracking algorithm to track the results of comparison at the same frame, and each column represents a tracking algorithm, from left to right the tracking algorithm followed by Mean-shift, SVM and our tracking method.

Figure 5. Sample results on occlusion1 sequence.

Fig.5 and Fig.6 show the tracking results using different tracking methods when the objects undergo severe occlusion on the Occlusion1 and Occlusion2 sequence. From Fig.5 we can see our tracking method performs well when the objects undergo severe occlusion, while the other methods including the Mean-shift and SVM methods completely fail to track the objects in these frames. From Fig.6 we can also see our tracking method performs well when the objects undergo severe occlusion, while the Mean-shift method fails to track the face at frames #140, #265 and #590. The main reasons are followings: (1) our method constructs a classifier by sampling positive and negative samples and then to find the best candidate that has the largest response using SVM classifier. (2)The proposed method integrates weighted multi-instance sampling method, which can consider the sample importance by the different weights.

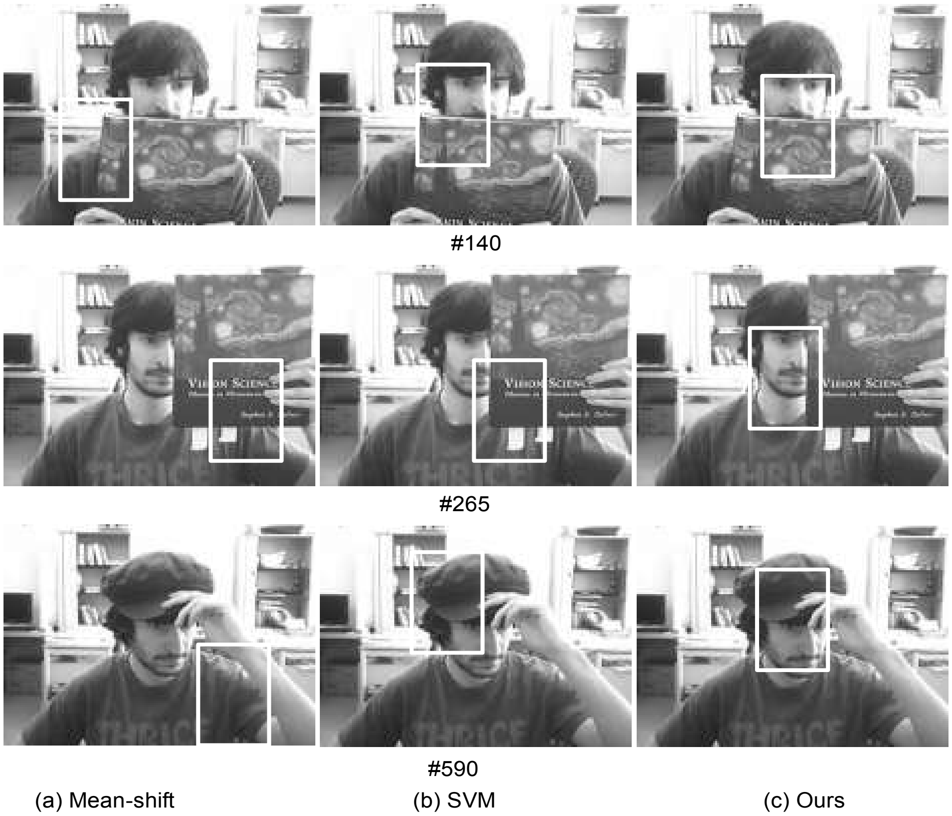


Figure 6. Sample results on Occlusion2 sequence.

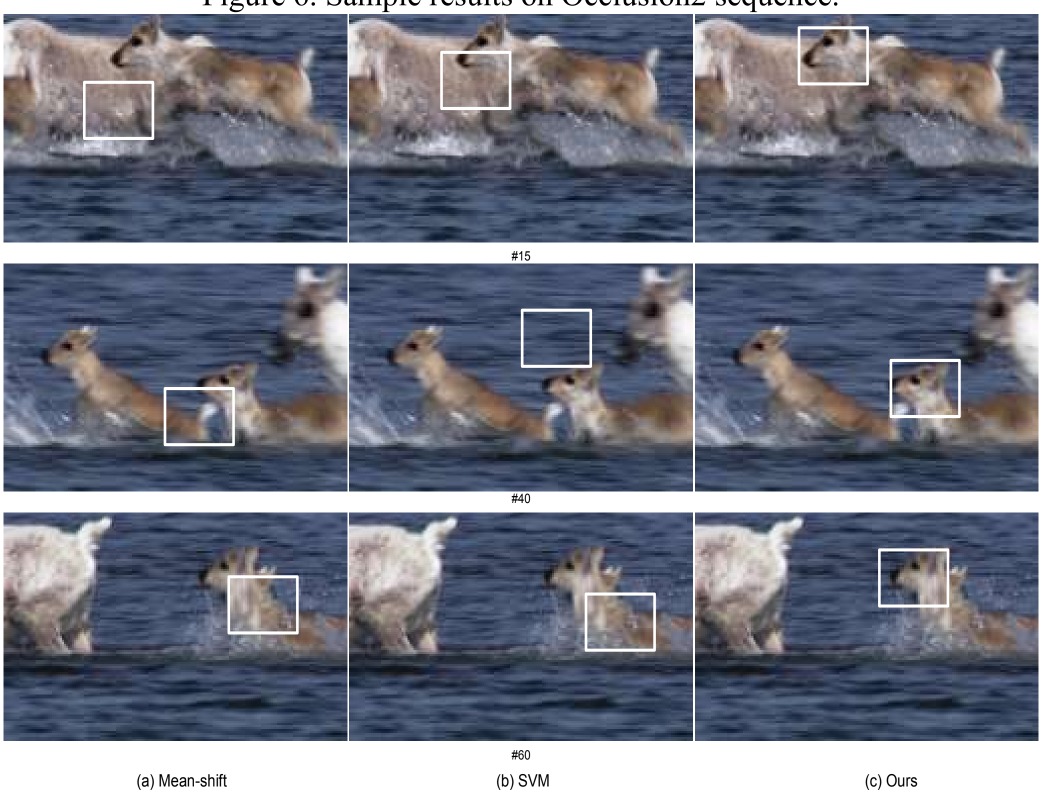


Figure 7. Sample results on deer sequence.

Figure 8. Sample results on David in door sequence.

Fig.7 and Fig.8 show the tracking results using different tracking methods when the objects undergo fast motion on the Deer sequence. From Fig.4 we can see our tracking method performs well when the objects undergo fast motion, while the other methods including the Mean-shift and SVM methods completely fail to track the objects in these frames. We can see the also tracking performance form Fig.6.

Fig.7 shows the tracking results using different tracking methods when the objects undergo illumination variation on the DavidIndoor sequence. From Fig.8 we can see our tracking method performs well when the objects undergo illumination variation, the Mean-shift and SVM methods fail to track the face at frames #180, #370 and #450.

A series of experimental results show that the visual tracking method proposed in this chapter translation in the target object, the rotary motion or illumination changes have good performance. In addition, we also the importance of the experiment proved the updated eigen subspace tracking methods are not update feature subspace.

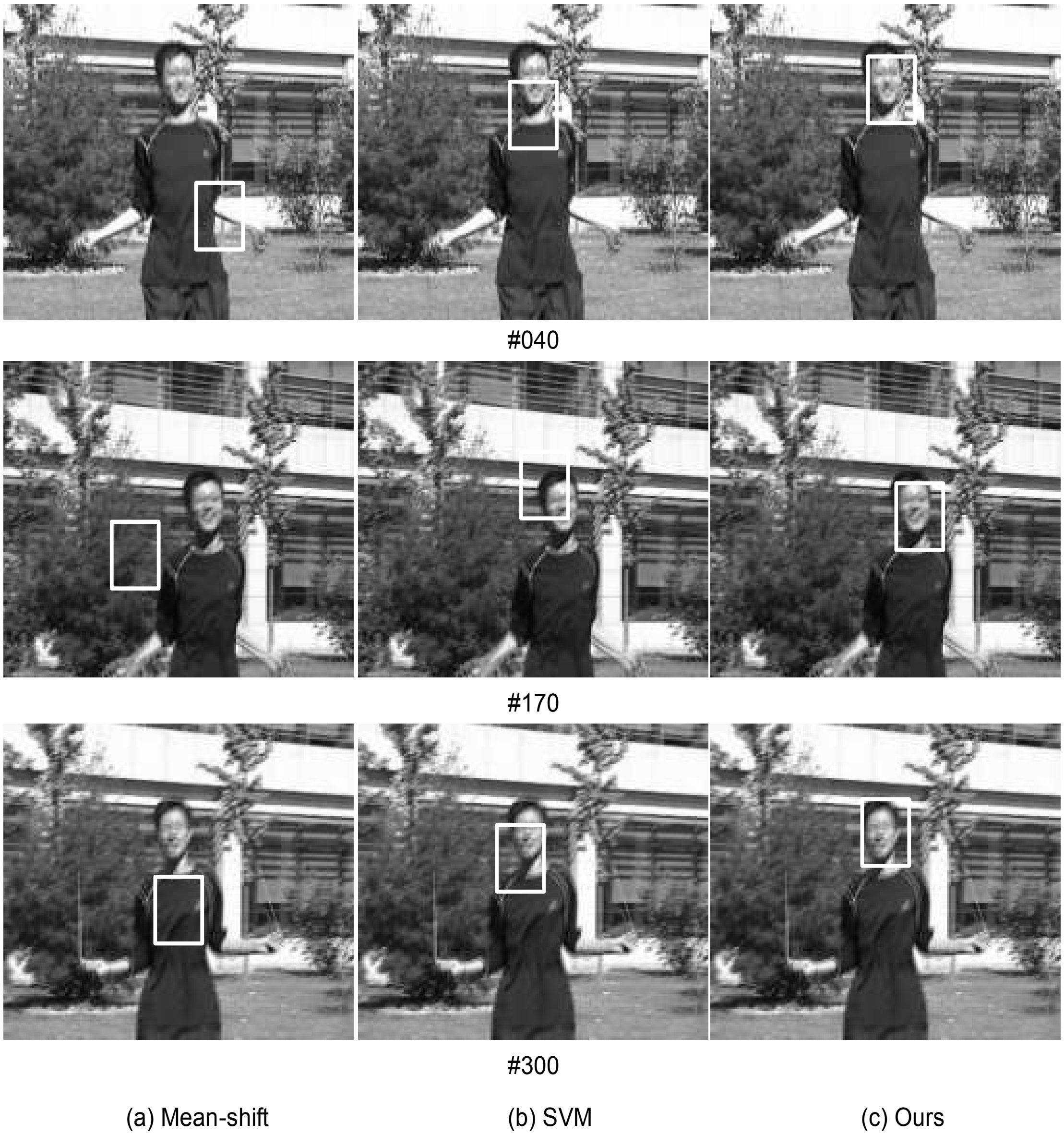


Figure 9. Sample results on jumping sequence.

V. CONCLUSION

It is a challenging task to develop effective and robust visual tracking method due to factors such as pose variation, illumination change, occlusion, and fast motion. In this paper, a novel tracking method based on SVM and weighted multi-sample sampling method. Our method constructs a classifier by sampling positive and negative samples and then to find the best candidate that has the largest response using SVM classifier. What's more, the proposed method integrates weighted multi-instance sampling method, which can consider the sample importance by the different weights. Experimental results on five challenge sequences show the accuracy and robustness of our proposed method.

VI. REFERENCES

[1] Isard M, Blake A. Condensation—conditional density propagation for visual tracking, International journal of computer vision, 1998, 29(1): 5-28.

[2] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: A review, Neurocomputing, 2011, 74(18): 3823-3831.

[3] Zhang S, Yao H, Sun X, et al. Sparse coding based visual tracking: Review and experimental comparison, Pattern Recognition, 2013, 46(7): 1772-1788.

[4] S.C.Mukhopadhyay, S. Deb Choudhury, T. Allsop, V. Kasturi and G. E. Norris, "Assessment of pelt quality in leather making using a novel non-invasive sensing approach", Journal of Biochemical and Biophysical methods, Elsevier, JBBM Vol. 70, pp. 809-815, 2008.

[5] Cannons K. A review of visual tracking, Dept. Comput. Sci. Eng., York Univ., Toronto, Canada, Tech. Rep. CSE-2008-07, 2008.

[6] Chang C, Ansari R. Kernel particle filter for visual tracking, Signal processing letters, IEEE, 2005, 12(3): 242-245.

[7] S.C.Mukhopadhyay, F.P.Dawson, M.Iwahara and S.Yamada, "A Novel Compact Magnetic Current Limiter for Three Phase Applications", IEEE Transactions on Magnetics, Vol. 36, No. 5, pp. 3568-3570, September 2000.

[8] Grabner H, Bischof H. On-line boosting and vision, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 1: 260-267.

[9] N.K. Suryadevara, S.C. Mukhopadhyay, R. Wang, R.K. Rayudu, Forecasting the behavior of an elderly using wireless sensors data in a smart home, Engineering Applications of Artificial Intelligence, Volume 26, Issue 10, November 2013, Pages 2641-2652, ISSN 0952-1976, http://dx.doi.org/10.1016/j.engappai.2013.08.004.

[10] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking [M]. Computer Vision–ECCV 2008. Springer Berlin Heidelberg, 2008: 234-247.

[11] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning [C]. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 983-990.

[12] Lowe D G. Distinctive image features from scale-invariant keypoints, International journal of computer vision, 2004, 60(2): 91-110.

[13] Stalder S, Grabner H, Van Gool L. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition[C]. Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on. IEEE, 2009: 1409-1416.

[14] N.K.Suryadevara, A. Gaddam, R.K.Rayudu and S.C. Mukhopadhyay, "Wireless Sensors Network based safe Home to care Elderly People: Behaviour Detection", Sens. Actuators A: Phys. (2012), doi:10.1016/j.sna.2012.03.020, Volume 186, 2012, pp. 277 – 283.

[15] Cortes C, Vapnik V. Support-vector networks, Machine learning, 1995, 20(3): 273-297.

[16] G. Sen Gupta, S.C. Mukhopadhyay, Michael Sutherland and Serge Demidenko, Wireless Sensor Network for Selective Activity Monitoring in a home for the Elderly, Proceedings of 2007 IEEE IMTC conference, Warsaw, Poland, (6 pages).

[17] Dalai N,Triggs B_ Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[18] M.Iwahara, S.C.Mukhopadhyay, S.Yamada and F.P.Dawson, "Development of Passive Fault Current Limiter in Parallel Biasing Mode", IEEE Transactions on Magnetics, Vol. 35, No. 5, pp 3523-3525, September 1999.

[19] Dalai N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance, Computer Vision-ECCV 2006. Springer Berlin Heidelberg, 2006: 428-441.

[20] Chen Q, Georganas N D, Petriu E M., Real-time vision-based hand gesture recognition using haar-like features, IEEE Instrumentation and Measurement Technology Conference Proceedings, 2007, IMTC 2007, 2007, pp. 1-6.

[21] Xiang Y, Su G., Multi-parts and multi-feature fusion in face verification, Computer Vision and Pattern Recognition Workshops, 2008. CVPRW 08. IEEE Computer Society Conference on IEEE, 2008: 1-6.

[22] G.Sengupta, T.A.Win, C.Messom, S.Demidenko and S.C.Mukhopadhyay, "Defect analysis of grit-blasted or spray printed surface using vision sensing technique", Proceedings of Image and Vision Computing NZ, Nov. 26-28, 2003, Palmerston North, pp. 18-23.

[23] Chen J, He Y, Wang J. Multi-feature fusion based fast video flame detection, Building and Environment, 2010, 45(5): 1113-1122.

[24] Khairul Anam, Adel Al Jumaily, Yashar Maali, Index Finger Motion Recognition Using Self-Advise Support Vector Machine, International Journal on Smart Sensing and Intelligent Systems, Vol. 7, no.2, pp. 644 – 657, 2014.

[25] G. Sen Gupta, S.C. Mukhopadhyay and M Finnie, Wi-Fi Based Control of a Robotic Arm with Remote Vision, Proceedings of 2009 IEEE I2MTC Conference, Singapore, May 5-7, 2009, pp. 557-562.

[26] Nicola Ivan Giannoccaro, Luigi Spedicato, Aime Lay-Ekuakille, A Robotic Arm to Sort Different Types of Ball Bearings from the Knowledge Discovered by Size Measurements of Image Regions and RFID Support, International Journal on Smart Sensing and Intelligent Systems, Vol. 7, No. 2, pp. 674 – 700, 2014.