ROBUST VISUAL TRACKING VIA MULTI-VIEW DISCRIMINANT BASED SPARSE REPRESENTATION

Bin Kang[1], *Dong Liang*[2], *Suofei Zhang*[1]

1. College of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, China
2. College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT

In traditional sparse representation based visual tracking, the particles are densely sampled, the appearance of some candidates may be very similar, hence the particle observations can be divided into disjointed groups. Existing methods only exploit the group similarity in a certain feature space. In this paper we propose a multi-view discriminant based multi-task sparse representation method to exploit the group similarity in a multi-feature space. The proposed method can discriminate the reliability of observation groups and achieve a proper multi-view fusion by using a multi-view discriminant matrix to project multi-feature observation groups into a common subspace. Experiment results show that our method can achieve a better tracking performance than state-of-the-art tracking methods do.

Index Terms— visual tracking, multi-view learning, sparse representation, dual group structure

1. INTRODUCTION

Visual tracking is of practical importance for many computer vision applications such as video surveillance, human-computer interaction and vehicle navigation etc. Since the target appearance can change dramatically under occlusion, background clutter or illumination change etc. The design of the observation model becomes the core of visual tracking. The state-of-the-art observation models in visual tracking can be broadly categorized into online classifier based models [1], template based models [2], subspace learning based models [3], sparse representation based models [4] and so on [5].

Among those algorithms, the sparse representation based visual tracking has been the centre of the research. Using sparse representation model for visual tracking was first proposed by Mei in [6]. This algorithm has a high computational complexity because it estimates the sparse representations of different particle observations separately. Although lots of works [7, 8, 9] have been done to improve the performance of Mei's algorithm, those trackers may drift away from the target in long term video sequences because they only use pixel intensity to model the target appearance. The pixel intensity is robust to particle occlusion but sensitive to the shape deformation of moving target and illumination change. In computer vision, multi-view refers to different feature subsets used to represent particular characteristics of an object. Based on this concept, Hong *et.al* [10] proposed a multi-view based algorithm for visual tracking, in which the target appearances in different feature space are linear combined by using a least absolute deviation based multi-task sparse representation model. The multi-task sparse representation model was derived based on the assumption that all the features can work well in visual tracking. It may not be valid in those video sequences with severe occlusion because some feature observations, such as texture, are prone to be disturbed by occlusion or video noise. In such case, the particle observations would not be well represented by their corresponding feature dictionary, it may cause a high reconstruction error. Fusing such feature observations with large reconstruction error may degrade the tracking performance.

Recently, Li *et.al* [11] proposed a group sparse coding based method for robust visual tracking, in which the author claimed that since the particles are densely sampled, the appearance of some candidates may be very similar, and therefore particle observations can be divided into disjointed groups. Exploiting the group similarity can reveal the potential commonality shared by the similar samples. Motivated by Li's work, in this paper, we first propose a multi-view discriminant based multi-task sparse representation method for robust visual tracking, in which we do not estimate the sparse representation of multi-view observation groups directly, instead, we estimate the sparse representation of multi-view group projections. Those group projections are obtained by using a learned multi-view discriminant matrix to project multi-feature observation groups into a common subspace. Introducing multi-view projection group into sparse repre-

Thanks to NUPT program NY 216023, National Key R. D. Program 2017YFB0802300, NSFC 61601223, NSF of Jiang Su Province BK20150756, and Post-doctoral Science Foundation of China 2015M580427 for funding.

sentation has two advantages: 1) It is difficult to exploit the group relationship in different feature space. The multi-view projections enable us exploit the group similarity in a common subspace. In this way, we can properly fuse multi-view observation groups to achieve robust sparse representation. 2) Due to the complementary of multi-view projections, we avoid introducing a large number of trivial templates in the target template matrix. Secondly, we propose an alternating algorithm to solve the optimization problem in the proposed sparse representation method.

2. MULTI-VIEW DISCRIMINANT BASED MULTI-TASK SPARSE REPRESENTATION

In this section, we first give a brief review of the traditional multi-task sparse learning based tracker, and then introduce the proposed sparse representation method in detail.

The traditional sparse representation based tracker is based on a particle filter framework [7], which mainly consists of two parts: the first part is to obtain particle observations using a particle filter; the second part calculates the posterior probability of each particle observation for the current frame. In the particle filter, the state vector of moving target at time t is denoted as $\mathbf{x}^t \in R^h$, and the observations of the state vector from time 1 to t is denoted as $\mathcal{Y}^t = \{\mathbf{y}^1, \mathbf{y}^2, ..., \mathbf{y}^t\}$. Using the Bayes rule, the posterior probability $p(\mathbf{x}^t|\mathcal{Y}^t)$ is calculated as $p(\mathbf{x}^t|\mathcal{Y}^t) \propto p(\mathbf{y}^t|\mathbf{x}^t) \int [p(\mathbf{x}^t|\mathbf{x}^{t-1}) p(\mathbf{x}^{t-1}|\mathcal{Y}^{t-1})] d\mathbf{x}^{t-1}$, where $p(\mathbf{y}^t|\mathbf{x}^t)$ is the observation likelihood and $p(\mathbf{x}^t|\mathbf{x}^{t-1})$ denotes the motion model. As it is very difficult to calculate $p(\mathbf{x}^t|\mathcal{Y}^t)$ directly using the aforementioned formula, the posterior probability is instead approximated by $p(\mathbf{x}^t|\mathcal{Y}^t) = \sum_{j=1}^{n} \omega_j^t \delta(\mathbf{x}^t - \mathbf{x}_j^t)$, where δ is the Dirac measure, $\mathbf{x}_j^t$ is the j-th sampled particle at time t, and ω_j^t is the particle importance weight, which is updated by $\omega_j^t = \omega_j^{t-1} p(\mathbf{y}^t|\mathbf{x}_j^t)$. The key to computing ω_j^t is to obtain $p(\mathbf{y}^t|\mathbf{x}_j^t)$. A traditional method to estimate $p(\mathbf{y}^t|\mathbf{x}_j^t)$ involves the use of a multi-task sparse learning method to jointly estimate the sparse representations of different particle observations. The multi-task sparse representation method is formulated as follows

$$\min_{\mathbf{\Theta}} \lambda \sum_{j=1}^{n} \|\boldsymbol{\theta}_j\|_1 + \frac{1}{2}\|\mathbf{Y} - \mathbf{A}\mathbf{\Theta}\|_F^2 \tag{1}$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]$ is the particle observation matrix at time t. $\mathbf{A}$ is the target template matrix, which is composed of target template vectors and a set of trivial templates, $\mathbf{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_n]$ is the sparse representation matrix of $\mathbf{Y}$. The main limitation of (1) is that it only use pixel intensity to model the target appearance, which is insufficient to deal with all kinds of appearance variations.

In this paper, we propose a multi-view discriminant based multi-task sparse representation method, which is shown as follows

$$\begin{aligned} \min_{\mathbf{\Theta},\mathbf{P}} \sum_{m=1}^{c} \sum_{i=1}^{k} \|\mathbf{P}_i^T \mathbf{Y}_i^m - \mathbf{P}_i^T \mathbf{A}_i \boldsymbol{\theta}_i^m\|_F^2 + \lambda_1 \|\mathbf{\Theta}\|_{2,1} \\ + \lambda_2 Tr(\mathbf{P}^T(\mathbf{S} - \mathbf{D})\mathbf{P}) \end{aligned} \tag{2}$$

where $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_k]$ and $\mathbf{P}_i$ denotes the learned discriminant matrix for the particle observations in the i-th feature space, $\mathbf{S}$ is the within-class scatter matrix in the common space and $\mathbf{D}$ is the between-class scatter matrix, both $\mathbf{S}$ and $\mathbf{D}$ are calculated through using the MvDA method [12], $\mathbf{Y}_i^m \in R^{u \times r}$ means the m-th particle observation group in the i-th feature space, which contains r particle observation vectors. $\mathbf{A}_i$ is the target template matrix in the i-th feature space, which can be updated in a manner similar to that in [13] to capture the variation of target appearance. $\mathbf{\Theta} = [\mathbf{\Theta}_1, \mathbf{\Theta}_2, ..., \mathbf{\Theta}_k]$, where $\mathbf{\Theta}_i = [\boldsymbol{\theta}_i^1, \boldsymbol{\theta}_i^2, ..., \boldsymbol{\theta}_i^c]$ and $\boldsymbol{\theta}_i^m$ $(i = 1, 2, ..., k, m = 1, 2, ..., c)$ denotes the sparse representation result of the m-th observation group projections in the i-th feature space. Problem (2) integrates multi-view learning [12] and sparse representation into a unified optimization model, in which we learn matrix $\mathbf{P}_i$ through solving $\min_{\mathbf{P}} Tr(\mathbf{P}^T(\mathbf{S} - \mathbf{D})\mathbf{P})$ to make $\mathbf{P}_i^T \mathbf{Y}_i^m$ more group-discriminative in the common space. In this case we can discriminate the reliability of different group projections. The within-group similarity in the observation groups can be enhanced by using the discriminant matrix. This means that the disturbance in different observation groups can be minimized. It is conductive to achieve a proper multi-view fusion. The sparse representation method in [11] is actually the special case of problem (2) because [11] only exploits the dual group structure in one feature space, our method considers the complementary of multi-feature and exploits the dual group structure in a multi-feature space.

Suppose $\mathbf{\Theta}$ is obtained through solving problem (2). We first find the group with minimum error, then the observation likelihood $p(\mathbf{y}^t|\mathbf{x}_j^t)$ in our tracking algorithm is calculated by

$$p(\mathbf{y}^t|\mathbf{x}_j^t) = \frac{1}{\Gamma} exp(-\alpha \sum_{i=1}^{k} \|\mathbf{P}_i^T [\mathbf{Y}_i^m]_j - \mathbf{P}_i^T \mathbf{A}_i [\boldsymbol{\theta}_i^m]_j\|_2^2) \tag{3}$$

where $[\mathbf{Y}_i^m]_j$ $(j = 1, 2, ..., r)$ means the j-th particle observation vector in the group matrix $\mathbf{Y}_i^m$ and $[\boldsymbol{\theta}_i^m]_j$ is the corresponding sparse representation result. The final optimal tracking result for the t-th frame is the weighted sum over all particles in a observation group with minimal sparse representation error, which is calculated as $\bar{\mathbf{x}}^t = \frac{\sum_{j=1}^{r} \omega_j^t \mathbf{x}_j^t}{\sum_{j=1}^{r} \omega_j^t}$, where ω_j^t is the particle weight of the j-th particle observation.

3. RECONSTRUCTION ALGORITHM

It is difficult to directly obtain an analytical solution to the objective function in (2). In this paper we propose an

alternating algorithm for the reconstruction of $\mathbf{\Theta}$ and $\mathbf{P}$. Each iteration of the alternating algorithm contains two steps: *P-step*, which aims at updating matrix $\mathbf{P}$, and *S-step*, which is to reconstruct the sparse representation matrix $\mathbf{\Theta}$

In *P-step*, we fix $\mathbf{\Theta}$ and update $\mathbf{P}$ by solving the following problem

$$\min_{\mathbf{P}} \sum_{m=1}^{c} \sum_{i=1}^{k} \|\mathbf{P}_i^T \mathbf{Y}_i^m - \mathbf{P}_i^T \mathbf{A}_i \boldsymbol{\theta}_i^m\|_F^2 + \lambda_2 Tr(\mathbf{P}^T(\mathbf{S}-\mathbf{D})\mathbf{P}) \tag{4}$$

To simplify problem (4), we can reformulate it as

$$\min_{\mathbf{P}} Tr(\mathbf{P}^T(\sum_{m=1}^{c}\sum_{i=1}^{k}(\mathbf{Q}_i^m)^T\mathbf{Q}_i^m)\mathbf{P}) + \lambda_2 Tr(\mathbf{P}^T(\mathbf{S}-\mathbf{D})\mathbf{P}) \tag{5}$$

where $\mathbf{Q}_i^m = (\mathbf{Y}_i^m - \mathbf{A}_i\boldsymbol{\theta}_i^m)^T$. Let $\mathbf{S}' = \lambda_2\mathbf{S} + \sum_{m=1}^{c}\sum_{i=1}^{k}(\mathbf{Q}_i^m)^T\mathbf{Q}_i^m$, then problem (5) becomes

$$\min_{\mathbf{P}} \lambda_2 Tr(\mathbf{P}^T(\mathbf{S}' - \lambda_2\mathbf{D})\mathbf{P}) \tag{6}$$

Problem (6) can be solved directly by setting its first derivation to zero, giving

$$\mathbf{D}^{-1}\mathbf{S}'\mathbf{P} = \lambda_2\mathbf{P} \tag{7}$$

The eigenvector matrix $\mathbf{P}^*$ with respect to $\mathbf{D}^{-1}\mathbf{S}'$ becomes the solution to problem (7)

In S-step, we fix $\mathbf{P}$ and update $\mathbf{\Theta}$ by solving following problem

$$\min_{\mathbf{\Theta}} \sum_{m=1}^{c}\sum_{i=1}^{k} \|\mathbf{P}_i^T\mathbf{Y}_i^m - \mathbf{P}_i^T\mathbf{A}_i\boldsymbol{\theta}_i^m\|_F^2 + \lambda_1\|\mathbf{\Theta}\|_{2,1} \tag{8}$$

In (8), let $F(\mathbf{\Theta}) = \sum_{m=1}^{c}\sum_{i=1}^{k}\|\mathbf{P}_i^T\mathbf{Y}_i^m - \mathbf{P}_i^T\mathbf{A}_i\boldsymbol{\theta}_i^m\|_F^2$, $\Omega(\mathbf{\Theta}) = \|\mathbf{\Theta}\|_{2,1}$, where $F(\cdot)$ is a differentiable convex function and $\Omega(\cdot)$ is a non-differentiable but convex function. Applying composite gradient mapping [14] to problem (8), we can obtain

$$\begin{aligned}\mathbf{\Theta}^{k+1} = &\min_{\mathbf{\Theta}} F(\mathbf{V}^k) + < \bigtriangledown F(\mathbf{V}^k), \mathbf{\Theta} > \\ &+ \frac{1}{2\eta}\|\mathbf{\Theta} - \mathbf{V}^k\|_F^2 + \lambda_1\Omega(\mathbf{\Theta})\end{aligned} \tag{9}$$

where η is a step-size parameter. Inspired by [15], the solution of (9) is given in closed-form by

$$\mathbf{\Theta}^{k+1/2} = \mathbf{V}^k - \eta \bigtriangledown F(\mathbf{V}^k) \tag{10}$$

where

$$\begin{aligned}&[\bigtriangledown F(\mathbf{V}^k)]_i^m = -(\mathbf{P}_i^T\mathbf{A}_i)^T(\mathbf{P}_i^T\mathbf{Y}_i^m - \mathbf{P}_i^T\mathbf{A}_i\mathbf{V}_i^m) \\ &i = 1, 2, ..., k \quad m = 1, 2, ..., c\end{aligned} \tag{11}$$

Based on (10), $\mathbf{\Theta}^{k+1}$ is finally updated by

$$[\mathbf{\Theta}^{k+1}](j,:) = [1 - \frac{\lambda_1\eta}{\|[\mathbf{\Theta}^{k+1/2}](j,:)\|_2}]_+[\mathbf{\Theta}^{k+1/2}](j,:) \tag{12}$$

where $[\mathbf{\Theta}^{k+1}](j,:)$ means the j-th row in matrix $\mathbf{\Theta}^{k+1}$. $[\cdot]_+$ is a scalar operator, suppose there is a scalar s, $[s]_+$ is defined as $[s]_+ = max\{0, s\}$. The complete algorithm proposed for solving problem (2) is summarized in the table below.

Algorithm 1 Alternating algorithm for solving problem (2)

Input: $\mathbf{Y}_i^m$, λ_1 and λ_2
Output: $\mathbf{\Theta}$, $\mathbf{P}$
Initialize: $k = 0$, $\mathbf{\Theta}^0 = \mathbf{V}^0 = \mathbf{0}$, $\mathbf{P} = \mathbf{0}$, $\alpha_0 = 1$
while not convergent **do**
P-step
1. Solving problem (6) for updating $\mathbf{P}^{k+1}$
S-step
2. Calculating $[\bigtriangledown F(\mathbf{V}^k)]_i^m$
3. $\mathbf{\Theta}^{k+1/2} = \mathbf{V}^k - \eta \bigtriangledown F(\mathbf{V}^k)$
4. $[\mathbf{\Theta}^{k+1}](j,:) = [1 - \frac{\lambda_1\eta}{\|[\mathbf{\Theta}^{k+1/2}](j,:)\|_2}]_+\mathbf{\Theta}^{k+1/2}](j,:)$
5. $\alpha_{k+1} = \frac{2}{k+3}$
6. $\mathbf{V}^{k+1} = \mathbf{\Theta}^{k+1} + \alpha_{k+1}(\frac{1-\alpha_k}{\alpha_k})(\mathbf{\Theta}^{k+1} - \mathbf{\Theta}^k)$
7. $k \longleftarrow k+1$
end while

4. EXPERIMENTS

In this section, we use the video sequences in CVPR2013 Visual Tracking Benchmark [16] to evaluate the performance of our proposed visual tracking algorithm. We compare the proposed tracking algorithm with seven state-of-the-art methods: IVT [3], CT [1], l1-APG [6], MTT [7], LRT [17], PLS [18] and CSK [19]. Since our method and the existing ones like the l1-APG, MTT and LRT are all particle filter based sparse representation algorithms, the particle number is set equally as 400.

Experiment setting: In our experiments, we use three complementary features to achieve visual tracking, namely intensity, edge with canny operator and texture with LBP [20]. The target template matrices are set as $\mathbf{A}_i \in R^{256\times 20}(i = 1, 2, 3)$, in which the particle observation size is 16×16, and the number of target templates is 20 (10 for foreground templates and 10 for background templates). The observation groups in different feature space are obtained by using the online K-means method [11]. The parameters in Algorithm 1 are set as $\eta = 0.01$, $\lambda_1 = 0.1\tau$ and $\lambda_2 = 0.1$.

Firstly, we evaluate the relationship between the group number and the tracking performance. We use online k-means [11] to achieve group division. From Fig.1 we could see that the proposed sparse representation method can give the best tracking performance when the particle observation is divided into 8 groups. If the group number is less than 8, some dissimilar particle samples may be involved in the particle observation groups and share a similar sparsity pattern with similar samples, it would degrade the tracking performance. If the group number is larger than 8, it can not make those

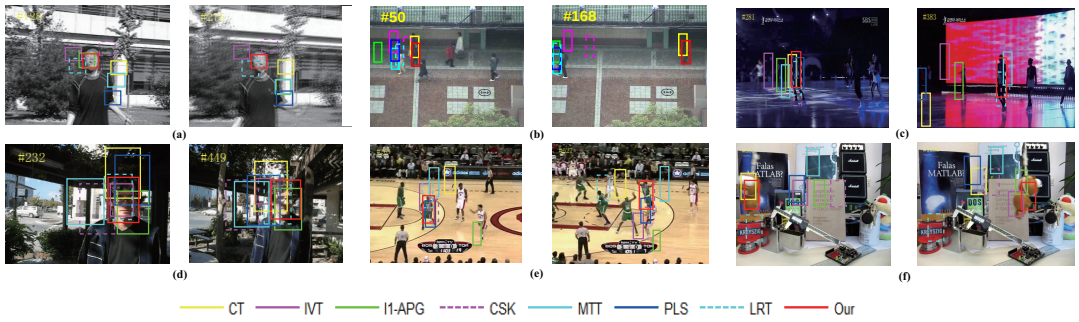

Fig. 2. Tracking results using 8 different methods on 6 challenging video sequences (a) jumping (b) subway (c) skating (d) trellis (e) basketball (f) lemming.

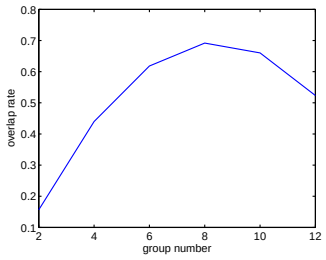

Fig. 1. Averaged overlap rate performance with varying candidate group numbers

similar particle samples to be grouped together, which would also reduce the tracking performance.

Next, we use 6 challenging video sequences to illustrate the qualitative tracking performance of our method. From Fig.2(a), we can see that the proposed tracking algorithm gives a significantly better tracking performance than other seven methods do. The challenges of Fig.2(b) are occlusion and background clutter. The tracking result indicates that our method can handel well background clutter. Fig.2 (c) and (d) have illumination changes, background clutters and non-rigid deformation. The proposed sparse representation model is able to explore the dual group structure in the multi-feature space, and thus can perform well in those two cases. The tests in Fig.2 (e) and (f) are very difficult because these is severe occlusion in there two video sequences. From Fig.2 (e) and (f) we can see that other methods fail to track the target whereas our method can still accurately track the moving target in the entire video sequence.

Finally, we give the quantitative evaluation of 15 tracking results(see Table.1 and 2). The quantitative visual tracking performance is evaluated by the center location error and overlap rate. The center location error is defined as the Euclidean distance between the central location of a bounding box and the labeled ground truth. The overlap rate is defined as $\frac{area(B_T \cap B_G)}{area(B_T \cup B_G)}$, where B_T and B_G are the tracking results of each frame and the corresponding ground truth, respectively.

5. CONCLUSION

In this paper, we have proposed a multi-view discriminant based multi-task sparse representation method for visual tracking. Our method has benefited from the use of a multi-view learning method that can exploit the group similarity between multi-view observatations in a common subspace. It has been shown through experiments that the proposed method can give a superior performance in challenging video sequences as compared to a number of known methods in literature.

Table 1. Average center location error over 15 video sequences. The best result is denoted as red

sequence	Proposed	IVT	CT	CSK	l1-APG	MMT	LRT	PLS
Car4	5.6	6.3	38.7	16.4	7.5	15.8	3.1	26.3
Faceocc1	3.6	7.7	12.3	5.2	10.5	5.3	6.5	7.0
Faceocc2	4.6	7.4	8.2	3.7	6.5	4.3	4.7	4.8
Basketball	6.1	44.2	64.3	15.1	97.4	80.6	59.1	45.5
Bolt	8.1	74.1	125.3	182.5	167.3	80.7	9.2	192.3
Carscale	3.4	18.8	26.6	19.5	10.7	17.0	12.8	31.4
Jumping	4.0	38.2	62.6	15.8	24.4	41.2	48.5	50.5
Lemming	9.8	42.5	28.7	31.6	49.6	74.0	85.6	78.2
David Indoor	7.6	15.6	9.9	18.2	27.5	11.6	21.7	40.1
Singer1	2.8	9.7	26.5	27.1	7.3	15.8	6.8	8.1
Skating	5.4	43.1	55.1	19.6	38.4	20.4	19.6	29.9
Subway	2.7	67.6	9.6	66.8	87.5	76.2	79.1	75.6
Sylvester	6.5	23.3	10.7	8.3	14.3	11.2	7.3	8.9
Trellis	6.1	23.8	21.3	41.7	14.9	28.2	9.8	23.9
Football	5.3	6.3	6.6	7.5	11.0	7.1	6.1	18.6

Table 2. Average overlap rate over 15 video sequences. The best result is denoted as red

sequence	Proposed	IVT	CT	CSK	l1-APG	MMT	LRT	PLS
Car4	0.78	0.76	0.27	0.36	0.61	0.51	0.80	0.34
Faceocc1	0.83	0.73	0.66	0.79	0.61	0.79	0.75	0.72
Faceocc2	0.72	0.69	0.60	0.78	0.61	0.72	0.71	0.71
Basketball	0.66	0.28	0.18	0.49	0.05	0.10	0.20	0.28
Bolt	0.69	0.04	0.02	0.01	0.03	0.02	0.65	0.02
Carscale	0.82	0.49	0.42	0.44	0.53	0.59	0.66	0.43
Jumping	0.63	0.21	0.04	0.17	0.36	0.20	0.12	0.06
Lemming	0.60	0.37	0.47	0.38	0.35	0.21	0.22	0.15
David Indoor	0.69	0.53	0.62	0.45	0.36	0.55	0.41	0.25
Singer1	0.79	0.51	0.35	0.29	0.55	0.42	0.61	0.55
Skating	0.61	0.15	0.24	0.43	0.18	0.45	0.48	0.43
Subway	0.76	0.16	0.54	0.19	0.02	0.17	0.15	0.16
Sylvester	0.65	0.36	0.55	0.63	0.43	0.52	0.62	0.60
Trellis	0.70	0.41	0.26	0.21	0.46	0.27	0.57	0.32
Football	0.72	0.65	0.55	0.54	0.49	0.57	0.67	0.47

6. REFERENCES

[1] K. Zhang, L. Zhang, and M.-H. Yang, "Fast compressive tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36 (10) pp. 2002-2015, 2014.

[2] Y. Wu, B. Shen, and H. Ling, "Visual tracking via online nonnegative matrix factorization," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24 (3) pp. 374-383, 2014.

[3] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," International Journal of Computer Version, vol. 77 (1) pp. 125-141, 2008.

[4] S. Zhang, H. Yao, X. Sun, and X. Lu, "Sparse coding base visual tracking: review and experimental comparison," Pattern Recognition, vol. 46 (7) pp. 1772-1788, 2013.

[5] X. Zhao, Y. Satoh, H. Takauji, and S. Kaneko, "Robust Tracking Using Particle Filter With a Hybrid Feature," IEICE Trans. on Information and Systems, vol.95-D(2), pp. 646-657, 2012.

[6] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33 (11) pp. 2259-2272, 2011.

[7] T. Zhang, G. Bernard, S. Liu, and A. Narendra, "Robust visual tracking via structured multi-task sparse learning," International Journal of Computer Vision, vol. 101 (2) pp. 367-383, 2013.

[8] Z. Xiao, H. Lu, and D. Wang, "L2-RLS-based object tracking," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24 (8) pp. 1301 - 1309, 2014.

[9] Shengping Zhang,Huiyu Zhou, Feng Jiang,and Xuelong Li, "Robust visual tracking using structurally random projection and weighted least squares," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25 (11) pp. 1749-1760, 2015.

[10] X. Mei, Z. Hong,, D. Prokhorov, and D. Tao, "Robust multitask multiview tracking in videos," IEEE Transactions on Neural Networks and Learning Systems, vol. 26 (11) pp. 2874-2890, 2015.

[11] F, Li, H. Lu, D. Wang, Y. Wu, and K. Zhang, "Dual group structured tracking," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26 (9) pp. 1697-1708, 2016.

[12] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38 (1) pp. 188-194, 2016.

[13] H. Li, C. Shen, and Q. Shi, "Real-time visual tracking using compressive sensing," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[14] X. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," IEEE Transactions on Image Processing, vol. 21 (10) pp. 4349-4360, 2012.

[15] M. Schmidt, E. Berg, M. Friedlander, and K. Murphy, "Optimizing costly functions with simple constraints: a limited-memory projected quasi-Newton algorithm," in Proc. Int. Conf. Artif. Intell. Stat., 2009.

[16] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark" in Proc. CVPR, 2013, pp. 2411-2418.

[17] T. Zhang, S. Liu, A. Narendra, M.-H. Yang, and G. Bernard, "Roubust visual tracking vis consistent low-rank sparse learning," International Journal of Computer Vision, vol. 111 (2) pp. 171-190, 2014.

[18] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Object tracking via partial least squares analysis," IEEE Transactions on Image Processing, vol. 21 (10) pp. 4454- 4465, 2012.

[19] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernel," in European Conference on Computer Vision (ECCV), 2012, pp. 702-715.

[20] T. Ojala, M. Pietikainen, and T. Maepaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24 (7) pp. 971-987, 2002.