VISUAL TRACKING VIA ROBUST MULTI-TASK MULTI-FEATURE JOINT SPARSE REPRESENTATION

Yong Wang, Xinbin Luo†, Shiqiang Hu†*

*Hisilicon Technologies
†School of Aeronautics and Astronautics, Shanghai Jiao Tong University

ABSTRACT

In this paper, we cast tracking as a novel multi-task learning problem and exploit various types of visual features. We use an on-line feature selection mechanism based on the two-class variance ratio measure, applied to log likelihood distributions computed with respect to a given feature from samples of object and background pixels. The proposed method is integrated in a particle filtering framework. We jointly consider the underlying relationship across different particles, and tackle it in a unified robust multi-task formulation. We show that the proposed formulation can be efficiently solved using the Alternating Direction Method of Multipliers (ADMM) with a small number of closed-form updates. Both the qualitative and quantitative results demonstrate the superior performance of the proposed approach compared to several state of-the-art trackers.

Index Terms**—** feature selection, multi-task learning, Alternating Direction Method of Multipliers

1. INTRODUCTION

Visual object tracking is one of the critical problems in computer vision. Various types of visual features including intensity [1-4], Haar feature [5, 6], color [7], texture [8] and superpixel [9] have been proposed.

A key issue addressed in our work is on-line, adaptive feature selection for tracking. Our insight is that the feature space that best distinguishes between foreground and background is the best feature space to be used for tracking. Meanwhile this choice of feature space will need to be continuously re-evaluated over time to adapt to changing appearances of the tracked object and scene background. Kalman filter algorithms are generally used to model object motion states in a visual tracking. It is employed to assign feature confidences, which ensures that the evolution of feature confidence is temporally consistency by exploiting the feature discriminative abilities in the current frame and feature confidences in the previous frames.

This work was jointly supported by the National Natural Science Foundation of China (No. 61374161) and China Aviation Science Foundation (No. 20142057006).

To overcome the problems mentioned above, we propose to employ other visual features such as color, edge, and texture to complement intensity in the appearance representation with a robust multi-task learning to solve the visual tracking problem. The workflow is shown in Fig.1. Within the proposed framework, the sparse representation for each feature is learned as a linear combination of atoms from an adaptive feature dictionary, i.e. each feature has its own sparse representation, which enables the tracker to capture different statistics carried by different features. To exploit the interdependencies shared between different particles, we impose the l1,2-norm group-sparsity on the representation matrix to learn the sparse representation jointly in a multi-task manner. To handle the outlier particles from particle sampling, we decompose the sparse representation into two collaborative parts, thereby enabling them to learn representative coefficients and detect outlier tasks simultaneously. An efficient Alternating Direction Method of Multipliers (ADMM) [10] scheme is employed to obtain the optimal solution via a sequence of closed-form updates.

Our contribution is three-fold: 1) we utilize multiple types of features in a sparse representation-based framework for tracking. Compared to previous related tracker, the new tracker is not only able to take advantage of the robustness to occlusion from sparse representation, but also introduces complementary multiple-feature representation for robust appearance modeling; 2) we treat every feature in each particle as an individual task and jointly consider the underlying relationship shared among different features and different particles in a multi-task learning framework; 3) to capture the outlier tasks that frequently emerge in the particle sampling process, we employ a robust multitask scheme by decomposing the coefficient matrix into two collaborative components.

2. OVERVIEW OF THE PROPOSED APPROACH

In our approach, four kinds of features are firstly extracted frame by frame. Then the feature evaluation is carried out to calculate the feature weight. Next, the particle filtering tracking method is employed for each feature. Finally, the tracking is completed by feature fusion.

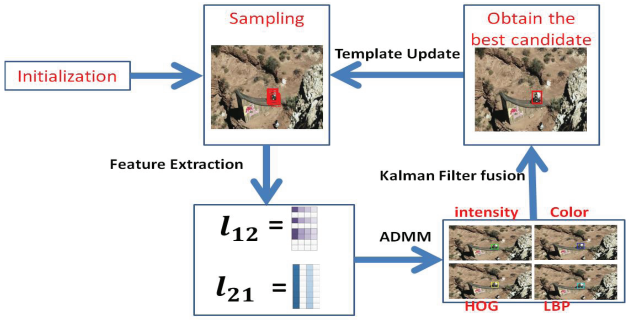

Fig. 1. The proposed tracking algorithm.

2.1. Feature Extraction

To take the advantage of complementary features, we employed four popular features: color histograms, intensity, histograms of oriented gradients (HOG) [11] and local binary patterns (LBP) [12]. HOG is a gradient-based feature that captures edge distribution of an object. LBP is powerful for representing object texture.

2.2. Object and Background Definition

We use a center-surround approach to sampling pixels from the object and the background. That is, a compact set of pixels (e.g. rectangle or ellipse) covering the object is chosen to represent the object pixels, while a larger surrounding ring of pixels is chosen to represent the background.

2.3. Discriminative Ability Calculation

We measure the discriminative ability [13, 14] of each feature in current frame by computing the likelihood ratio between the object feature and its corresponding background feature as follows:

$$\tilde{R_t^i} = max(0, min(1, log\frac{max(F_t^i(x,y),\delta)}{max(B_t^i(x,y),\delta)})), i = 1, \cdots, N \quad (1)$$

$$R_t^i = \frac{\tilde{R_t^i}}{\Sigma \tilde{R_t^i}} \quad (2)$$

Where $F_t^i(x,y)$ and $B_t^i(x,y)$ are the ith features of the object and the background at frame t respectively. δ is set as 0.005 empirically.

2.4. Object Tracking

We carry out the object tracking in particle filtering framework. The particle filtering can be divided into the prediction and the update steps:

$$p(ss_t|y_{1:t-1}) = \int p(ss_t|x_{t-1})p(ss_{t-1}|y_{1:t-1})dss_{t-1} \quad (3)$$

$$p(ss_t|y_{1:t}) = \frac{p(y_t|ss_t)p(ss_t|y_{1:t-1})}{p(y_t)|y_{1:t-1}} \quad (4)$$

where $ss_{1:t} = \{ss_1, ss_2, ..., ss_t\}$ denotes the state vectors up to time t and $y_{1:t} = \{y_1, y_2, ..., y_t\}$ are observations variables. $p(ss_t|ss_{t-1})$ is a dynamic model that describes the state transition, and $p(y_t|ss_t)$ is an observation model for each state. For each state ss_t. Thus, the optimal object state at time t can be determined by solving the maximum a posterior (MAP) problem.

Kalman filter affords a solution to the feature evaluation during a tracking process. In common sense, features of higher discriminative ability have larger confidences and vice versa. Therefore, we define the state of the Kalman filter for feature evaluation as the combination of the confidence R_t and its variation "velocity" δR_t of each feature. The measurement of the filter is the discriminative ability vector at frame t. Then the weight of the feature can be determined by the prediction equation and the measurement equation of a Kalman filter [14].

$$y = \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 \quad (5)$$

where $\alpha_i, i = 1, 2, 3, 4$ is the normalized Kalman filter prediction.

2.5. Two-norm based Representation and Decomposition

In our tracking model, the particles are represented by a linear combination of templates in dictionary D_t, i.e.,

$$X_t = D_t W_t + E_t \quad (6)$$

where W_t is a coefficient matrix, $W_t = [w_1, w_2, \cdots, w_n]_t, D_t = [d_1, d_2, \cdots, d_(N_T)]_t, N_T$ is the number of template. In (6), the D_t is decomposed into two part, row sparsity for capturing the shared features among all particles and column sparsity to identify the outlier noise simultaneously. Thus, the particle representation in our tracker becomes the following form:

$$X_t = D_t W_t + E_t = D_t(L_t + S_t) + E_t \quad (7)$$

Decomposing matrix W_t into two separate sparse matrixes (i.e., L_t and S_t) brings high robustness for the proposed tracker. More specifically, we formulate our tracking model as a group sparsity learning scheme:

$$min_{L_t,S_t} \parallel X_t - D_t(L_t + S_t) - E_t \parallel_F^2 + \lambda_1 \parallel L_t \parallel_{1,2} + \lambda_2 \parallel S_t \parallel_{2,1} + \lambda_3 \parallel E_t \parallel_{1,1}, W_t = L_t + S_t \quad (8)$$

where L_t is the row group sparsity component, S_t is the column group sparsity component and E_t is the noise. The three

Algorithm: group sparsity algorithm implemented by ADMM.

Input: X_t, D,

Initialize L, S, (here L and S represent L_t and S_t, respectively; t is omitted for clarity of the algorithm description in the following.)

k=1,

While stopping criterion is not met do

Solve the convex optimization problem:

$$L^{k+1} = \underset{L^k}{\text{minimize}} \frac{1}{2}\left\|(X_t - D^k S^k - E^k) - D^k L^k\right\|_F^2 + \lambda_1 \left\|Z_L^k\right\|_{1,2}$$

$$s.t.\ L^k = Z_L^k,$$

Perform the iterations of scaled ADMM algorithm:

$$L^{k+1} = ((D^k)^T D^k + \rho I)^{-1}[(D^k)^T (X_t - D^k S^k) + \rho(Z_L^k - u_L^k)],$$

$$Z_L^{k+1} = prox_{12}(Z_L^k),$$

$$u_L^{k+1} = u_L^k + L^{k+1} - Z_L^{k+1},$$

where $prox_{12}$ denotes proximal method for $\|*\|_{1,2}$ [16].

Solve the convex optimization problem with updated L^{k+1}:

$$S^{k+1} = \underset{S^k}{\text{minimize}} \frac{1}{2}\left\|(X_t - D^k L^{k+1} - E^k) - D^k S^k\right\|_F^2 + \lambda_2 \left\|Z_S^k\right\|_{2,1}$$

$$s.t.\ S^k = Z_S^k,$$

Perform the iterations of scaled ADMM algorithm:

$$S^{k+1} = ((D^k)^T D^k + \rho I)^{-1}[(D^k)^T (X_t - D^k L^{k+1}) + \rho(Z_S^k - u_S^k)],$$

$$Z_S^{k+1} = prox_{21}(Z_S^k),$$

$$u_S^{k+1} = u_S^k + S^{k+1} - Z_S^{k+1},$$

where $prox_{21}$ denotes proximal method for $\|*\|_{2,1}$ [16].

Solve the convex optimization problem with updated E^{k+1}:

$$E^{k+1} = \underset{E^k}{\text{minimize}} \frac{1}{2}\left\|(X_t - D^k L^{k+1} - D^k S^{k+1}) - E^k\right\|_F^2 + \lambda_3 \left\|E^k\right\|_{1,1}$$

$$s.t.\ E^k = Z_E^k,$$

The scaled form of ADMM consists of the following iteration.

$$E^{k+1} = \left((D_t^k)^T D_t^k + \rho I\right)^{-1} [X_t - D^k L^{k+1} - D^k S^{k+1} + \rho(Z_E^k - u_E^k)],$$

$$Z_E^{k+1} = prox_{11}(Z_E^k),$$

$$u_E^{k+1} = u_E^k + E^{k+1} - Z_E^{k+1},$$

where $prox_{11}$ denotes proximal method for $\|*\|_{1,1}$.

k = k+1;

end while

Output: solution $L_t = L^{k+1}$, $S_t = S^{k+1}$, $W_t = L_t + S_t$, to equation (8).

Fig. 2. The proposed tracking algorithm.

parameters λ_1, λ_2 and λ_3 balance reliable construction and joint sparsity, where λ_1 regulates the row group sparsity on L_t, λ_2 controls the column group sparsity on S_t and λ_3 is a regularization parameter which encourages the error matrix to be sparse.

2.6. Resolve Equation (8)

Equation (8) represents the key idea of our method. Resolving equation (8) is essential to efficiently compute matrix W_t. While there are several optimization algorithms, we choose the ADMM algorithm for its simplicity and efficiency. We summarize our group sparsity algorithm implemented by ADMM for resolving equation (8) in Fig.2. More detail can be found in [10, 15].

2.7. Template Update

To handle appearance variations, the target dictionary D_t is progressively updated similar to [1], and the templates are weighted in the course of tracking.

	CT	IVT	L1-APG	L1	L2-RLS	MIL	MTT-L01	MTT-L11	WMIL	Ours
Caviar	68.5046	85.6776	24.8298	19.0748	144.9574	69.7605	65.2356	103.1512	88.6514	4.8362
Caviar1	16.8755	33.3381	48.4095	106.8751	1.3008	87.2633	53.4084	101.8416	29.531	3.6022
Caviar2	63.167	13.3922	5.8703	24.6669	16.3851	22.6452	4.8235	10.4497	62.0663	6.0703
Cup	43.6092	1.6257	2.4408	2.9189	2.6194	40.9867	64.6413	159.8587	10.1017	3.1348
DavidIndoor	16.5161	67.5226	31.2135	86.6966	20.8459	23.1548	88.3262	19.5485	23.577	14.4856
Dog1	15.7522	45.5823	9.7048	52.7659	4.6743	16.8385	8.2277	9.7217	29.521	9.5283
Faceocc	18.8312	64.9672	24.8161	40.7633	20.7714	32.9755	33.3148	164.7155	49.0825	17.4954
Faceocc2	24.5455	63.7754	12.4189	41.2344	11.5001	21.4552	8.1904	29.6458	30.9168	14.9717
Human	3.4695	359.7149	1.921	125.4265	393.4827	5.8683	3.2527	92.2673	16.3524	5.4127
Jump	216.0381	234.5108	172.217	239.6197	234.287	219.8929	6.2736	121.3118	16.553	9.3426
Shirt	12.2288	111.6935	21.7862	45.5775	87.4274	22.4791	72.3955	205.4118	26.456	7.473
Shop	70.5524	7.8512	2.9598	3.9316	61.0245	15.3831	2.4224	23.0991	62.8289	6.8675
Ucsdpeds	5.3104	11.4702	1.7455	43.7165	62.797	10.9856	1.2932	4.6251	12.555	5.2814
Crossing	7.3319	78.4715	2.2177	2.7864	2.4461	5.8021	3.2711	42.5012	108.3541	2.8952
David2	80.5788	48.9736	2.6076	58.8368	1.9651	14.3728	1.4139	2.9057	11.106	4.4797
Freeman1	14.1436	74.6266	8.4877	57.1923	14.6514	14.7425	121.7953	111.3064	22.5456	14.106
Head_motion	15.2259	27.0431	9.437	8.6842	8.2647	9.8749	8.2459	9.4153	89.1344	8.7958
Jogging	107.8466	22.9099	85.1418	87.5305	96.0912	86.7589	128.4964	104.8154	95.3137	7.3243
Mhyang	31.5107	51.5048	3.6666	35.4632	9.7712	53.9352	4.4252	19.2103	43.3887	6.0334
Motocross-2	10.2428	42.9039	31.7723	54.4645	35.8466	21.4	63.9674	13.4824	65.1716	11.1312
Wpolarbear	13.4378	21.9712	9.4285	30.2203	12.7469	11.8766	12.092	19.6445	32.9759	7.0886
Xrocky	170.3048	160.0649	5.4093	97.6778	9.8632	168.8514	82.5297	204.0934	115.5571	9.4632

Fig. 3. The average tracking errors. The error is measured using the Euclidian distance of two center points from the ground truth. The last row is the average error for each tracker over all the test sequences.

3. EXPERIMENT

Performance of the proposed tracker is analyzed on 22 challenging video sequences and compared with seven state-of-the-art tracking works including the Incremental Visual Tracking (IVT) [22], L1 tracking (L1T) [1], L1-APG tracking [2], multi-task tracking (MTT-L01, MTT-L21) [3], Multiple Instance Learning tracking (MIL) [5], compressive tracking (CT) [6], Wacv12 [17], WMIL [21], LSST [19], L2-RLS [18]. The sequences include either a nonrigid object or an object that undergoes significant appearance changes. The tracker was implemented in Matlab and runs at approximately 1 frame per second on an Intel Core i5. The trackers are run 3 times and the average results are reported for each video clip. We would like to emphasize that all the parameters were kept constant for all experiments.

3.1. Quantitative Comparison

We evaluate the above-mentioned algorithms using the center location error as well as the overlapping rate [20], and the results are shown in Fig.3 and Fig.4. Overall, the proposed tracking algorithm achieves the best or second best results in most sequences in terms of both success rate and center location error.

3.2. Qualitative Comparison

Scale Change and Occlusion: Caviar2 and Shop contain significant occlusion and scale change. Our method tracks the target person well until the end of these sequences. In the

	CT	IVT	L1-APG	L1	L2-RLS	MIL	MTT-L01	MTT-L11	WMIL	Ours
Caviar	0.158	0.146	0.22	0.424	0.024	0.162	0.156	0.15	0.138	0.988
Caviar1	0.3927	0.3089	0.3037	0.301	1	0.0079	0.3037	0.2984	0.0288	0.9843
Caviar2	0.27	0.302	0.914	0.4	0.342	0.036	0.982	0.424	0.012	0.936
Cup	0.4587	1	0.9967	1	1	0.4455	0.4752	0.1617	0.7294	1
DavidIndoor	0.2359	0.2273	0.2078	0.2121	0.2273	0.0606	0.2857	0.3701	0.2143	0.6364
Dog1	0.6022	0.223	0.9993	0.5659	0.9993	0.5778	0.8652	0.7333	0.2222	0.7926
faceocc	0.5011	0.0609	0.2167	0.1907	0.3309	0.0011	0.0158	0.1151	0.0011	0.55
Faceocc2	0.5767	0.3877	0.4147	0.4442	0.7067	0.5607	0.9607	0.8613	0.5546	0.7914
Human	0.5049	0.0243	0.9078	0.3058	0.0218	0.4029	0.9976	0.2451	0.267	0.9345
Jump	0.1842	0.0219	0.3509	0.1096	0.136	0.1272	0.7588	0.3158	0.307	0.8377
Shirt	0.7939	0.0126	0.6036	0.3396	0.0053	0.7056	0.0053	0.0053	0.2818	0.9853
Shop	0.3393	0.4036	0.9768	0.9768	0.3643	0.3393	0.9929	0.3625	0.0125	0.8411
Ucsdpeds	0.5594	0.0383	1	0.341	0.023	0.0575	0.8352	0.4751	0.0038	0.6858
Crossing	0.85	0.025	0.8583	0.9583	0.2417	0.8167	1	0.424	0.0083	0.9583
David2	0.0019	0.0857	0.8361	0.257	0.9646	0.0037	1	0.1617	0.3259	0.9311
Freeman1	0.6055	0.0123	0.7393	0.1687	0.3129	0.0706	0.1442	0.3613	0.0828	0.6319
Head_motion	0.9489	0.6711	0.7545	0.9838	0.9728	1	0.983	0.2451	0.0694	0.997
Jogging	0.2248	0.2345	0.2248	0.0065	0.0423	0.2085	0.2248	0.3158	0.1889	0.8502
Mhyang	0.002	0.294	0.9913	0.5745	0.7517	0.002	1	0.0493	0	0.9456
Motocross-2	0.8696	0.3043	0.5217	0.2609	0.3913	0.6087	0.5652	0.0053	0.0435	1
Wpolarbear	0.6173	0.2588	0.7305	0.4286	0.6092	0.7008	0.6846	0.5445	0.1078	0.8464
Xrocky	0.0533	0.1036	0.5	0.0414	0.7456	0.0562	0.5207	0.1183	0.003	0.6479

Fig. 4. Average overlap rate. The best three results are shown in red, blue, and green fonts.

Fig. 5. Shows screenshots of some tracking results.

Caviar2, the WMIL method loses the target when another person occludes the target subject, and the CT also misses the target when occlusion happens.

Scale and Illumination Change. For the Crossing sequence, when the light changes drastically, WMIL and IVT fail to track the object reliably. The proposed method is robust to scale and illumination changes as object appearance can be modeled well by multiple feature fusion and the representation model is used to separate the inliers and noise.

Abrupt motion and pose variation. The Shirt sequence contains object with abrupt motion and pose variation. In the sequence, our method is capable of tracking the target for the entire sequence whereas other methods gradually drift away. The reason that our method performs well can be explained as follows. First, our tracker is able to reduce drifts with the use of multiple features. Second, the proposed algorithm uses representation model that account for large and drastic appearance change. In addition, our method employs ADMM algorithm to solve the model.

In plane rotation and Background clutters. The target objects in the David2 sequence undergo in plane rotation and background clutters. Our method performs well throughout the sequence.

Scale change. The Ucsdpeds sequence contains significant scale change. In the sequence, two people walk away from the camera. The $L1$ method loses track of the target from the start. All the other methods successfully track the target but the MTT-L01, MTT-L02, L1-APG, and our method achieve higher overlap scores.

4. CONCLUSION

In this paper, we have presented a robust multi-feature fusion and multi-task joint sparse learning method for particle filtering based tracking. By appropriately introducing the $l_{1,2}$ norm regularization, the method not only exploits the underlying relationship shared by different features and different particles, but also captures the frequently emerging outlier tasks. We implemented our method using four types of complementary features, i.e. intensity, color histogram, HOG and LBP. A feature selection criterion is employed to continually evaluate and update the features used for tracking. The ADMM algorithm is used to solve the formulation. Kalman filter based fusion scheme is used to fuse the features. The experimental results demonstrate that the proposed method is capable of taking advantage of features and correctly handling the outlier tasks. Compared to popular trackers, our tracker demonstrates superior performance. Moreover, the proposed method can potentially be extended to handle data obtained from sensors other than cameras.

5. REFERENCES

[1] Mei, X., Ling, H., Robust visual tracking and vehicle classification via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(11), pp. 2259C2272, (2011)

[2] C. Bao, Y. Wu, H. Ling, and H. Ji, Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Rhode Island, (2012)

[3] Zhang, T., Ghanem, B., Liu, S., Ahuja, N., Robust visual tracking via multi-task sparse learning. In IEEE conference on computer vision and pattern recognition (pp. 1C8), (2012)

[4] Zhang, T., Ghanem, B., Liu, S., Ahuja, N, Low-rank sparse learning for robust visual tracking. In Computer VisionCECCV (pp. 470-484). Springer Berlin Heidelberg, (2012)

[5] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie, Robust Object Tracking with Online Multiple Instance Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8), 1619-1632, (2011)

[6] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, Real-Time Compressive Tracking. Proceedings of European Conference on Computer Vision, vol. 3, pp. 864-877, Florence, Italy, October, (2012)

[7] M. Isard and A. Blake, Condensation conditional density propagation for visual tracking. IJCV, 29(1):5C28, (1998)

[8] Z. Kalal, K. Mikolajczyk, and J. Matas, Tracking-Learning-Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7): 1409-1422, (2011)

[9] S. Wang, H. Lu, F. Yang, and M.-H. Yang, Superpixel tracking, in Proc. IEEE Int. Conf. Comput. Vision, Nov. 2011, pp. 1323C1330.

[10] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J, Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn., 3(1):1C122, (2011)

[11] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of IEEE Conference on CVPR, 2005, pp. 1063C6919.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. TPAMI, 24(7):971C987, 2002.

[13] R. T. Collins, Y. Liu, and M. Leordeanu. Online selection of discriminative tracking features. TPAMI, 27(10):1631C1643, 2005.

[14] Zhenjun Han, Qixiang Ye, Jianbin Jiao, Combined feature evaluation for adaptive visual object tracking, Computer Vision and Image Understanding, Volume 115, Issue 1, January 2011, Pages 69-80

[15] Yong Wang, Shiqiang Hu, and Shandong Wu, ”Visual tracking based on group sparsity learning,” Machine Vision and Applications, pp: 1-13, 2014.

[16] X. Chen, W. Pan, J. Kwok, and J. Carbonell, Accelerated gradient method for multi-task sparse learning problem. In IEEE international conference on data mining (pp. 746C751), (2009)

[17] Qing Wang, Feng Chen, Wenli Xu, Ming-Hsuan Yang. Online Discriminative Object Tracking with Local Sparse Representation. IEEE Workshop on the Applications of Computer Vision, 425-432, 2012.

[18] Ziyang Xiao, Huchuan Lu, Dong Wang: L2-RLS-Based Object Tracking. IEEE Trans. Circuits Syst. Video Techn. 24(8): 1301-1309 (2014)

[19] Dong Wang, Huchuan Lu, Ming-Hsuan Yang: Least Soft-Threshold Squares Tracking. CVPR 2013: 2371-2378

[20] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results, (2010)

[21] Zhang K, Song H. Real-time visual tracking via online weighted multiple instance learning [J]. Pattern Recognition, 2013, 46(1): 397-411.

[22] Ross, D., Lim, J., Lin, R.S., Yang, M.H., Incremental learning for robust visual tracking. International Journal of Computer Vision, 77(1), 125C141, (2008)