ROBUST OBJECT TRACKING VIA MULTI-TASK BASED COLLABORATIVE MODEL

Yong Wang, Xinbin Luo†, Shiqiang Hu‡*

*School of Electrical and Computer Engineering, University of Ottawa.
†School of Electronic information and Electrical Engineering, Shanghai Jiao Tong University
‡School of Aeronautics and Astronautics, Shanghai Jiao Tong University

ABSTRACT

This paper presents a robust object tracking algorithm using a collaborative model. Under the framework of particle filtering, we develop a multi-task learning based generative and discriminative classifier model. In the generative model, we propose a histogram-based subspace learning method that takes advantage of adaptive template update. In the discriminative model, we introduce an effective method to compute the confidence value that assigns more weights to the foreground than the background. A decomposition model is employed to take the outliers of each particle into consideration. The alternating direction method of multipliers (ADMM) algorithm guarantees the optimization problem can be solved robustly and accurately. Qualitative and quantitative comparison with ten state-of-the-art methods demonstrates the effectiveness and efficiency of our method in handling various challenges during tracking.

Index Terms— Collaborative model, Alternating direction method of multipliers, Multi-task learning

1. INTRODUCTION

Visual object tracking has been extensively studied in computer vision due to its importance in applications ranging from robotics, surveillance, augmented reality to traffic monitoring ([1-9]). Appearance model plays a key factor in a tracking system. An object can be represented by different features such as intensity [3-5], superpixels [2], Haar-like features [8] [9] and color [7]. Since the simplicity and efficiency, in this paper, we use intensity values to represent objects.

Tracking algorithms can be categorized into two classes based on their representation schemes: generative [3-6], [10], [12-15] and discriminative models [1], [8], [9], [11], [16].

The generative method which learns the appearance of an object has been exploited to handle variations of the target. Though much demonstrated success of generative tracking algorithms, two problems remain to be solved. First, most tracking algorithms assume that the target appearance does not change much during this period. However, the drift problem is likely to occur if the appearance of the target changes dramatically. Second, these generative algorithms do not use the background information which is likely to improve tracking stability and accuracy. The discriminative tracker aims to find a decision boundary that can best separate the target from the background. A major problem of this model is to rely heavily on training sample selection.

Generally speaking, when there is less variability in the tracked object, generative trackers tend to yield more accurate results than discriminative trackers because generative methods typically use richer features. However, in more complicated environments, discriminative trackers are often more robust than generative trackers because discriminative trackers use negative samples to avoid the drifting problem. A natural attempt is to combine the two approaches to a hybrid approach, as in [17, 19, 29]. The collaborative multi-task sparse representation is employed in [29]. The generative appearance representation is based on [5].bAnd approximate proximal gradient (APG) is used to solve the representation. The experimental results show the collaborative model improves the tracking accuracy. The appearance models of the work mentioned above are mainly based on generative methods. Thus in this paper, we make use of the generative model to account for appearance change and the discriminative classifier to effectively separate the foreground target from the background.

Through the above analysis, we propose a multi-task sparse tracking framework which makes use of the generative model to account for appearance change and the discriminative classifier to effectively separate the foreground target from the background. Our contributions can be summed up in the following three aspects:

First, we propose an innovative decomposition formulation to model the appearance of the object. The decomposition model considers the target, background and noise. Alternating direction method of multipliers (ADMM) algorithm guarantees the convergence of the formulation.

Second, we propose a simple yet useful framework to integrate the generative and discriminative model. While generative trackers usually produce more accurate results under less complex environments due to the richer image represen-

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

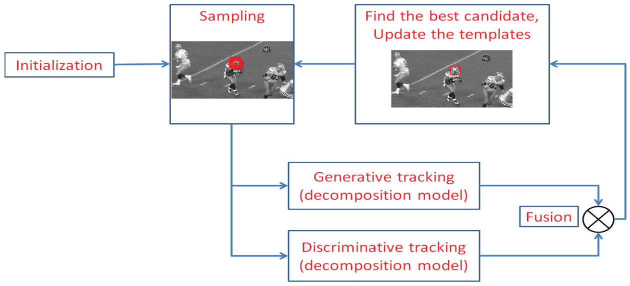

Fig. 1. The proposed tracking framework.

tations used, discriminative trackers are more robust against strong occlusion and variations since they explicitly take the background into consideration. The complex tracking scenario requires the trackers must possess the property of handle variety situations.

Third, extensive experiments against the state-of-the-art methods are carried out to demonstrate the effectiveness of our method.

2. PROPOSED ALGORITHM

In this section, we present the proposed algorithm in details. We first describe the particle filtering. Then the generative method with holistic feature is illustrated. Next the discriminative method with foreground and background information is exploited. Following is the collaborative model. Finally the update scheme is presented. Fig. 1 illustrates the overall flows of the proposed tracking framework.

2.1. Particle Filtering Tracking Framework

In this section, we give a detailed description of our particle filtering based tracking method. Particles are sampled at and around the previous object location to predict the state ss_t of the target at time t. The state transition function $P(ss_t|y_t)$ is modeled by an affine motion model with a diagonal Gaussian distribution. The observation model $P(y_t|ss_t^i)$ reflects the similarity between an observed image region y_t corresponding to a particle ss_t^i and the templates of the current dictionary. The particle that maximizes this function is selected to be the tracked target at each time instance.

2.2. Generative Method

At each time t the tracking target and all candidates are modeled with k PCA basis vectors and an error term as:

$$Y_t = D_t W_t + E_t, \tag{1}$$

where $Y_t \in R^{d*N}$, $Y_t = [y_1, y_2, \cdots, y_N]$ is the observation vector, N is the number of obervations, $D_t \in R^{d*k}$ denotes a matrix of PCA basis vectors (d represents feature

Input: X_t, D_t,

Initialize L, S, E, (here L, S and E represent L_t, S_t and E_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:

$$L^{k+1} = \underset{L^k}{\text{minimize}} \frac{1}{2} \|(Y_t - D_t^k S^k - E_t^k) - D_t^k L^k\|_F^2 + \lambda_1 \|Z_L^k\|_{1,\infty}$$

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

Perform the iterations of scaled ADMM algorithm:

$$L^{k+1} = \left((D_t^k)^T D_t^k + \rho I\right)^{-1} \left[(D_t^k)^T (Y_t - D_t^k S^k - E_t^k) + \rho(Z_L^k - u_L^k)\right],$$

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

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

where $prox_{1inf}$ denotes proximal operator of $\|*\|_{1,\infty}$, $\rho > 0$.

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

$$S^{k+1} = \underset{S^k}{\text{minimize}} \frac{1}{2} \|(Y_t - D_t^k L^{k+1} - E_t^k) - D_t^k S^k\|_F^2 + \lambda_2 \|Z_S^k\|_{2,1}$$

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

Perform the iterations of scaled ADMM algorithm:

$$S^{k+1} = \left((D_t^k)^T D_t^k + \rho I\right)^{-1} \left[(D_t^k)^T (Y_t - D_t^k L^{k+1} - E_t^k) + \rho(Z_S^k - u_S^k)\right],$$

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

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

where $prox_{21}$ denotes proximal operator of $\|*\|_{2,1}$.

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

$$E^{k+1} = \underset{E_t}{\text{minimize}} \frac{1}{2} \|(Y_t - D_t^k L^{k+1} - D_t^k S^{k+1}) - E_t^k\|_F^2 + \lambda_3 \|E_t^k\|_{1,1}$$

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

The scaled form of ADMM consists of the following iteration.

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

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

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

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

k = k+1;

end while

Output: solution $L_t = L^{k+1}$, $S_t = S^{k+1}$, $W_t = L_t + S_t$, $E_t = E^{k+1}$, to equation (2).

Fig. 2. Implementation of the proposed algorithm.

dimension and k the number of PCA basis), $W_t \in R^{k*N}$ denotes the corresponding coding vectors (target coefficients), and $E_t \in R^{d*N}$ represents the error term. We use the affine transformation to model the object motion between two consecutive frames.

Recently the $l_{1,\infty}$ norm has been proposed for joint regularization. Essentially, this type of regularization aims at learning a set of joint sparse models. The $l_{1,\infty}$ norm is a matrix norm that penalizes the sum of maximum absolute values of each row. The $l_{2,1}$ norm regularization is employed to encourage column sparsity. Thus, based on the group sparsity learning principle, our tracker works by searching a row sparse matrix L_t and a column sparse matrix S_t, regulating the common and different information across the templates, respectively. The $l_{1,1}$ norm is used to regularize the noise term.

Now equation (1) can be re-written as the following form:

$$Y_t = D_t W_t + E_t = D_t(L_t + S_t) + E_t \tag{2}$$

The above equation can be transformed to the following equation:

$$\begin{aligned} min_{L_t,S_t,E_t} ||Y_t - D_t(L_t + S_t) - E_t||_F^2 + \lambda_1 ||L_t||_{1,\infty} \\ + \lambda_2 ||S_t||_{2,1} + \lambda_3 ||E_t||_{1,1}, W_t = L_t + S_t \end{aligned} \tag{3}$$

where the three parameters λ_1, λ_2 and λ_3 are weights that quantify the trade-off between different terms. Fig. 2 shows the decomposition model computed by the ADMM algorithm. The detail of the ADMM algorithm can be found in [20, 21]. We summarize our group sparsity generative tracking algorithm in Fig. 3.

Algorithm1: group sparsity generative tracking algorithm

Input: Current frame at t.
Dictionary template D.
All n particles ss_{t-1}.

1. Generate n particles ss_t within the particle filtering framework.
2. Compute imaging feature for each of the n particles and then form matrix Y_t.
3. Obtain group sparse representation $W_t = L_t + S_t$ by solving equation (1).
4. Calculate reconstruction error $\Delta r_i = \|y_t^i - Dw_i - e_i\|_2$, $i = 1, \ldots, n$.
5. Calculate $p_g(y_t^i|ss_t) = exp(-\Delta r_i^2)$ for each particle.
6. Select the particle with the highest value of $p_g(y_t^i|ss_t)$ as the current tracking result y_t^i.

Output: Tracked target y_t^i.
Current state ss_t.

Fig. 3. The proposed generative tracking algorithm.

2.3. Discriminative Method

The logistic model is employed in the discriminative part in [28]. Our discriminative model is still use the multi-task learning. The framework is based on [25]. However, we use the decomposition form to replace the Laplacian form in [25] since our decomposition form is more suitable to represent the target. Therefore, it is easier to separate the target from the background. A discriminative model for object representation is presented and the background information around the target is taken into consideration.

The positive and negative template sets are defined as $Ts_{pos} = [t_1, t_2, \cdots, t_p]$ and $Ts_{neg} = [t_{p+1}, t_{p+2}, \cdots, t_{p+n}]$ respectively, where p and n denote the number of positive and negative templates. With these assumptions, the problem formulation is equivalent to a multi-task learning problem as equation (3). The coefficient matrix $W_t = [w_1, \cdots, w_{p+n}]_t$ is is computed by the ADMM algorithm as mentioned above.

Each column of W denotes the coefficients of a certain template decomposed by all candidates.

$$f_i = [W_{i1}, \cdots, W_{ip}, W_{i(p+1)}, \cdots, W_{i(p+n)}]^T \quad (4)$$

where W_{ij} is the element in the i-th row and the j-th column of W. Thus, we define a similarity map: $F = [f_1, \cdots, f_N] = W^T$, where each column is a discriminative feature of a candidate, indicating its similarity levels to p positive templates and n negative templates. Next, two sub similarity maps of $Fpos = W_{pos}$ and $Fneg = W_{neg}$ are defined.

The difference between the j-th candidate y_j and the i-th template Ts_i is defined as:

$$WD_{ij} \propto = exp(-||Ts_i - y_j||_2^2) \quad (5)$$

A candidate share higher similarity with foreground templates should corresponds to a small weight while those share lower similarity with templates should correspond to a large weight. In the following, we define two sub weight map WD_{pos} and WD_{neg} for the positive samples and negative samples respectively. Two weighted maps is defined as $\hat{F}_{pos} = WD_{pos} \odot F_{pos}$ and $\hat{F}_{neg} = WD_{neg} \odot F_{neg}$ respectively, where $\odot$ is the Hadamard product (element-wise product).

Then the first l largest F_{pos} and F_{neg} is defined as $s_{i-pos} = L(\hat{F}_{i-pos}, 1) + \cdots + L(\hat{F}_{i-pos}, l)$ and $s_{i-neg} = L(\hat{F}_{i-neg}, 1) + \cdots + L(\hat{F}_{i-neg}, l)$ respectively, where i denote the i-th candidate. The discriminative score for the i-th candidate is formulated by

$$s_i = s_{i-pos} - s_{i-neg} \quad (6)$$

A larger score indicates the candidate is more likely to be the object and vice versa. The likelihood of the observation y_t^i being the target at state ss_t can be defined in the following:

$$p_d(y_t^i|ss_t) \propto s_i \quad (7)$$

2.4. Collaborative Model

The two models can be unified within the particle filter framework. The likelihood function of the i-th candidate is constructed by

$$p(y_t^i|ss_t) = p_g p_d \quad (8)$$

and the tracking result is the candidate with the highest probability.

2.5. Update Scheme

For our tracking model, we update the negative templates every several frames (5 in our experiments) from image regions away (e.g., more than 8 pixels) from the current tracking result. The positive templates remain the same in the entire sequence.

3. EXPERIMENT

Performance of the proposed tracker has analyzed on twenty six challenging video sequences and compared with ten state-of-the-art tracking works including the Incremental Visual Tracking (IVT) [10], L1 tracking (L1T) [3], L1-APG tracking [4], multi-task tracking [5] Multiple Instance Learning tracking (MIL) [8], compressive tracking (CT) [9], Wacv12 [18], WMIL [1], LSST [24], L2-RLS [23]. 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 2 frames per second on an Intel Core i5. We will make the source code available to the public.

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 are kept constant for all experiments.

Two evaluation criteria are used in the experiments [22]: center location error and tracking success rate, both computed against manually labeled ground truth.

	CT	L2-RLS	LSST	Wmil	IVT	L1T	L1-APG	Mil	Mtt-l01	Wacv12	Our
Bird_2	58.835	99.855	88.8215	19.888	103.5217	45.6218	39.554	176.162	73.0801	72.9999	26.7162
Cup	43.609	2.619	1.665	10.102	1.6257	2.9189	2.4408	40.9154	64.6413	109.2349	1.622
David	43.5624	137.3245	168.8119	29.06428	27.789	109.51	26.6646	43.4271	27.9621	89.6949	8.413
DavidIndoor	27.7209	26.4424	186.8211	30.1069	69.5102	94.1248	36.527	53.9578	74.2636	91.7653	6.6625
Dog1	15.7522	4.67428	10.4541	29.521	45.5823	52.7659	9.7048	16.3329	8.2277	13.0824	11.6438
Dollar	14.2953	6.498	64.909	73.4557	43.7373	66.4509	66.8405	22.4236	6.37995	34.7709	8.13686
DudekData	96.7557	17.5674	78.8483	39.8296	58.5268	103.8284	29.0199	22.441	14.79	104.348	13.6206
Faceocc2	24.5455	11.5001	14.2359	30.9168	63.7754	41.2344	12.4189	48.4556	8.19044	23.8746	11.048
Football	13.2756	76.0107	41.653	17.0534	223.595	70.0796	16.879	15.8858	14.6457	26.5027	11.6288
Juice	6.7165	6.15447	1.6828	10.4687	69.1284	33.5829	0.9835	12.2762	3.4975	18.393	1.4058
Jump	216.038	234.287	155.697	16.553	234.511	239.6197	172.217	228.918	6.2736	13.7984	7.992
Shop	70.5524	61.0245	67.5826	62.8289	7.8512	3.9316	2.9598	121.302	2.4224	28.5934	2.4755
Singer	19.2334	27.082	3.8067	18.061	47.1161	78.7356	5.098	19.1367	51.9938	7.3943	6.1762
Skating1	140.075	83.226	132.926	171.692	159.207	141.53	38.5085	111.291	326.504	128.553	24.344
Ucsdpeds	5.3104	62.797	1.706	12.555	11.4702	43.7165	1.7455	4.3716	1.2932	8.4437	4.2519
Crossing	7.33192	2.4461	2.354	108.354	78.4715	2.7864	2.2177	10.7198	3.2711	62.2966	2.8927
David2	80.5788	1.965	5.9766	11.106	48.9736	58.8368	2.6076	56.5513	1.4139	32.5989	2.1731
Fish	12.3659	38.1686	6.763	52.5336	33.1363	90.0293	19.052	12.7524	39.8556	54.8687	4.5227
Freeman1	14.1436	14.6514	30.3282	22.5456	74.6266	57.1923	8.4877	21.3484	121.795	113.904	8.6507
Head_motion	15.2259	8.2647	9.9178	89.1344	27.04	8.6842	9.437	61.3272	8.2459	19.223	9.544
Mhyang	31.5107	9.77122	2.22448	43.3887	51.5048	35.4632	3.66661	22.1423	4.4252	15.0193	2.4885
Motocross_2	10.2428	35.8466	46.3636	65.1716	42.9039	54.4645	31.7723	55.2349	63.9674	10.6539	7.27752
Chasing	12.8025	5.8737	4.1249	9.3662	38.6676	44.4327	4.9867	12.0286	6.7806	24.6546	6.47376
Wsurfing	10.6356	2.027	3.1946	14.5308	76.054	1.6554	1.5356	14.8301	1.3556	8.2617	1.6159
Xpets2000	11.3008	9.4355	18.8506	16.8258	8.4054	11.022	11.3609	15.5153	4.8607	6.3738	8.4686

Fig. 4. The average tracking errors.

	CT	L2_RLS	LSST	wmil	IVT	L1T	L1-APG	Mil	Mtt_l01	Wacv12	Our2
Bird_2	0.4444	0.0404	0.1515	0.5051	0.061	0.4747	0.2828	0.0404	0.1515	0.101	0.404
Cup	0.4587	1	1	0.7294	1	1	0.9967	0.4521	0.4752	0.1089	1
David	0.0303	0.0801	0.0693	0.041	0.1082	0.0195	0.0931	0.041	0.0866	0.04978	0.7965
DavidIndoor	0.4048	0.04545	0.02597	0.3788	0.0519	0.1017	0.04545	0.0714	0.0974	0.001299	0.9649
Dog1	0.6022	0.99926	0.8541	0.2222	0.2230	0.5659	0.9993	0.6148	0.8652	0.7689	0.8563
Dollar	0.9694	0.9817	0.3914	0.1927	0.2661	0.3914	0.3945	0.6177	0.9878	0.2202	0.9817
DudekData	0.2541	0.8279	0.7153	0.4882	0.5109	0.5179	0.593	0.9362	0.934	0.2437	0.9354
Faceocc2	0.5767	0.7067	0.535	0.5546	0.3877	0.4442	0.4147	0.1828	0.9607	0.4908	0.8712
Football	0.7265	0.4448	0.5387	0.5525	0.1436	0.1768	0.6298	0.6657	0.7569	0.4972	0.59392
Juice	0.4702	0.8861	0.9975	0.4579	0.3490	0.6238	1	0.25495	1	0.1559	1
Jump	0.1842	0.13596	0.3465	0.307	0.0219	0.1096	0.3509	0.0439	0.7588	0.5044	1
Shop	0.3393	0.3643	0.3571	0.01250	0.4036	0.9768	0.9768	0.00179	0.99286	0.10000	0.9375
Singer	0.2906	0.0256	1	0.2593	0.3447	0.2536	0.4359	0.1026	0.3476	0.359	1
Skating1	0.1475	0.04250	0.11000	0.04250	0.06000	0.1575	0.45500	0.00750	0.12250	0.16500	0.61250
Ucsdpeds	0.5594	0.02299	0.9464	0.0038	0.038	0.341	1	0.5670	0.8352	0.00383	0.5785
Cliff_dive_1	0.7632	0.5921	0.5658	0.6316	0.5658	0.4737	0.4342	0.0921	0.6184	0.7105	0.6316
Crossing	0.8500	0.2417	0.3083	0.00833	0.02500	0.9583	0.85833	0.60000	1	0.1250	0.9000
David2	0.0019	0.9646	0.6946	0.3259	0.0857	0.257	0.8361	0.00372	1	0.01117	1
Fish	0.9139	0.0651	0.3088	0.0357	0.2164	0.0336	0.0735	0.8151	0.04202	0.1282	1
Freeman1	0.2055	0.3129	0.6104	0.0828	0.01227	0.16878	0.2393	0.0675	0.1442	0.1779	0.7055
Head_motion	0.9489	0.9728	0.9987	0.0694	0.67106	0.9838	0.7545	0.00766	0.9830	0.1583	0.9800
Mhyang	0.0020	0.7517	1	0.00067	0.29396	0.5745	0.9913	0.4604	1	0.9013	1
Motocross_2	0.8696	0.3913	0.3913	0.0435	0.3043	0.2609	0.5217	0.1304	0.5652	0.8696	1
Chasing	0.1783	0.6717	0.8983	0.7400	0.0667	0.62	0.6400	0.5967	0.7050	0.3117	0.8683
Wsurfing	0.9326	0.9965	1	0.3972	0.0426	1	1	0.2908	1	0.6773	1
Xpets2000	0.3595	0.527	0.5405	0.1351	0.6946	0.4108	0.1189	0.0649	0.7081	0.8541	0.6892

Fig. 5. The average overlap rate.

3.1. Quantitative Comparison

The average tracking errors and average overlapping errors are presented in Fig. 4 and Fig. 5 respectively. The best three results are shown in red, green, and blue fonts respectively. 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

Fig. 6 shows screenshots of some tracking results.

Scale change. In the Dog1 sequence, the scale of the Dog changes significantly. Our representation model represents the appearance of the dog accurately. Thus our method can track the target throughout this sequence and achieve higher overlap scores.

Abrupt motion and illumination. In the fish sequence, the target undergo abrupt motion in cluttered backgrounds and large illumination change. Our generative model tracks the target accurately at most of the time. Therefor our collaborative model successfully tracks the target for the entire sequence.

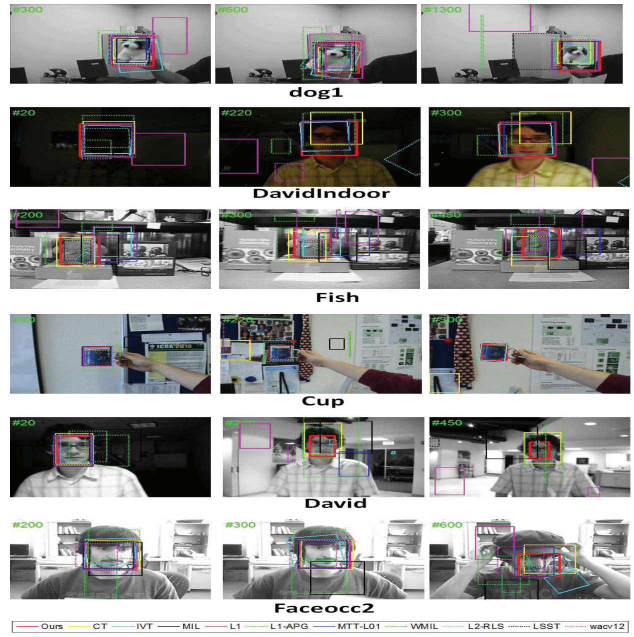

Fig. 6. Shows screenshots of some tracking results.

Illumination, pose variation and scale changes. The david and davidIndoor sequences contain large illumination, pose variation and scale changes. The generative model tracks the target when illumination happened. The update scheme guarantees the templates gradually follow the change of the target.

Background Clutter: The cup sequence includes scenes with cluttered background. The decomposition model reconstructs the appearance of the cup accurately. Thus it helps the discriminative model to locate the target.

Occlusion: In the Faceocc2 sequence, the target is severely occluded by a book. Our decomposition model alleviates the effect of the occlusion. Thus the generative model and the discriminative model locate the targets successfully. The MIL method loses track of the face and starts to follow the book when partial occlusion occurs. Other trackers are able track the face accurately except the IVT, L1 and Wacv12 methods.

3.3. Conclusion

In this paper, we propose a robust tracking method based on the collaboration of generative and discriminative method. Holistic templates are incorporated into generative method to deal with cluttered and complex background. Positive and negative templates are adapted to discriminative method to separate foreground and background. The multi-task sparse representation scheme learns the candidates jointly. Extensive experimental results on visual tracking have demonstrated the effectiveness and robustness of the proposed method.

4. REFERENCES

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

[2] S. Wang, H. Lu, F. Yang, and M.-H. Yang, Superpixel tracking, in Proc. IEEE Int. Conf. Comput. Vision, Nov. 2011, pp. 1323-1330.

[3] Mei, X., Ling, H., Robust visual tracking and vehicle classification via sparse representation. IEEE Transactions on PAMI, 33(11), 2259-2272, (2011)

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

[5] Zhang, T., Ghanem, B., Liu, S., Ahuja, N., Robust visual tracking via multi-task sparse learning. In IEEE conference on CVPR (pp. 1-8), (2012)

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

[7] D. Comaniciu, V. R. Member, and P. Meer, Kernel-based object tracking. IEEE Transactions on PAMI, 25(5):564-575, (2003)

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

[9] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, Real-Time Compressive Tracking. Proceedings of ECCV, vol. 3, pp. 864-877, Florence, Italy, October, (2012)

[10] Ross, D., Lim, J., Lin, R.S., Yang, M.H., Incremental learning for robust visual tracking. International Journal of Computer Vision, 77(1), 125-141, (2008)

[11] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang, Real-Time Object Tracking via Online Discriminative Feature Selection, IEEE Transactions on Image Processing, 22(12), pp.4664-4677, (2013)

[12] Dong Wang, Huchuan Lu, and Ming-Hsuan Yang, Online Object Tracking with Sparse Prototypes. IEEE Transactions on Image Processing, 22(1), 314-325, (2013)

[13] Y. Wu, B. Shen, and H. Ling, Visual Tracking via Online Non-negative Matrix Factorization. IEEE Trans. on Circuits and Systems for Video Technology, in press

[14] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, Efficient Minimum Error Bounded Particle Resampling L1 Tracker with Occlusion Detection. IEEE Trans. on Image Processing, 2014

[15] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, Robust tracking using local sparse appearance model and k-selection. In CVPR, (2011)

[16] H. Grabner and H. Bischof, On-line boosting and vision. In CVPR, (2006)

[17] T. B. Dinh and G. G. Medioni, Co-training framework of generative and discriminative trackers with partial occlusion handling. In Proceedings of IEEE Workshop on Applications of Computer Vision, pages 642-649, (2011)

[18] 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.

[19] Wei Zhong, Huchuan Lu, and Ming-Hsuan Yang, Robust Object Tracking via Sparse Collaborative Appearance Model, IEEE Transactions on Image Processing (TIP), vol. 23, no. 5, pp. 2356-2368, 2014.

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

[21] 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): 1-122, (2011)

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

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

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

[25] Bohan Zhuang, Huchuan Lu, Ziyang Xiao, Dong Wang: Visual Tracking via Discriminative Sparse Similarity Map. IEEE Transactions on Image Processing 23(4): 1872-1881 (2014)

[26] Zhang, T, Ghanem, B., Liu, S, Ahuja, N: Robust Visual Tracking via Structured Multi-Task Sparse Learning. IJCV, (2013)

[27] Xue Mei, Zhibin Hong,Danil Prokhorov, Dacheng Tao: Robust Multitask Multiview Tracking in Videos. IEEE Transactions on Neural Networks and Learning Systems. 26(11) (2015)

[28] Xu Cheng et al, Object tracking via collaborative multi-task learning and appearance model updating. Applied Soft Computing (2015)