ROBUST OBJECT TRACKING BASED ON DISCRIMINATIVE ANALYSIS AND LOCAL SPARSE REPRESENTATION

Peng Tian[1], JiangHua Lv[1,]*
1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China
* Corresponding Author: jhlv@nlsde.buaa.edu.cn

ABSTRACT

To improve robustness in cases of partial occlusion, deformation and rotation in visual tracking, local similarity measurements are usually used. However, this method have drawbacks under complex backgrounds. For example, the method only consider the traditional similarity measurements of objects and templates, results in the matching errors are prone to lead to the failure of tracking. In this paper, we proposes a object tracking algorithm based on measurements of the local discriminative similarities. This new method have advantages as following: firstly, both the similarities and the discrimination are considered; Secondly, the discriminative weight learning of the local region is carried out to improve the accuracy of fragment measurement; At last, an effective and efficient tracker is designed based on the difference analysis and a simple update manner within the particle filter framework. Experimental results show that the proposed algorithm achieves better performance than traditional competing methods

Index Terms—*Object tracking; local sparse representation; similarity measurement; discriminative analyses ; discriminative weight*

1. INTRODUCTION

Object tracking is an important field in the computer vision research and has many applications in daily life [1] such as human-computer interaction, video surveillance, robotic vision and others [2]. However, affected by many interference factors such as deformation, scales change, varying illumination, fast motion, as well as the dynamic and complex environments, the object always undergoes large and unpredictablechanges in their visual appearance. Therefore, similarity measurements are often inaccurate. Object tracking is still a challenging task [3].

The representations of the objects are an important factor in the similarity measurement. According to the object area, the method can be divided into two parts: the global and local representation. The global measurement method could overcome the scale, translation and rotation without distortion [4], however, in practical applications, the change of light, deformation, occlusion and other issues can not be effectively resolved [6,7,8]. Comparatively, the local representation of the object provides the spatial information of the feature. And it is robust to the problems of deformation and occlusion [9,10,11]. In recent years, the similarity measure method based on local region has drawn much attention.

Adam et al. [11] proposed an appearance modeling for tracking based on local fragments, which divides the target into non-overlapping fragment. By combining the similarity among the histograms of the different fragment, the object can be tracked. To improve the accuracy of tracking in further, He [12] uses the EMD distance to compute the distance between the two local sensitive histogram of the fragment. The locally orderless matching (LOM), which was proposed according to the EMD distance measurement for the fragment in the LOT algorithm [13], using the LOM matching to match the most similar fragments and to improve the accuracy of the object tracking. However, in this algorithm, the tracking accuracy is reduced because the different fragments are equally treated and the difference of the fragments are ignored. To solve this problem, Wang Dong [14] proposed a method of the local weighted cosine measurement similarity, it works through the cosine measurement analysis of different noise of the fragment measurement, to form the new method of the local cosine measurement. It relieve local occlusion later by online weight learning.

The above methods acquire similarity analysis of the object under the local similarity measurement methods. However, in tracking processes, the accuracy of the complex changes, such as illumination, posture and part occlusion, is declined and this measurement does not change adaptively in changing cases. This measurement actually relies only on the similarity measures. It is so easily to disturbed by the noise. Essentially, the object tracking is to look for the object which have been tracked and is different from the background.Therefore, both the "object and object" similarity and the "object and background" discrimination are important in tracking. So, the object tracking, as the "object and background" classification problem, also need to consider the discrimination between the object and the background [15]. Consequently, this paper proposes a similarity measure based on local discriminant analysis to measure the
similarity between the target and the sample. The algorithm based on the sparse representation and the discriminative analysis, achieves the local similarity measurement by combining the similarity and the discrimination, it further

This work is supported by the National Natural Science Foundation Project of China (Grant No. 61300007).

studies the differential weight to improve the accuracy of the fragment measurement and make the tracking more accurate.

2. THE SIMILARITY MEASURE OF LOCAL DISCRIMINATIVE STRUCTURE

Traditional similarity measurement usually use the global description method and only consider the similarity of the object-object, thus they can not effectively distinguish the changes of the feature with the changes of the scene and state, because there are full of noise, background diversity and other factors so that the difference between the sample and the background is ignored in the tracking process. In this paper, we propose the Similarity Measure of Local Sparse Structure based on the similarity of the target and the discrimination of the background to promote the accuracy of the similarity measurement.

A. Local sparse histogram representation

In consideration of wide application of sparse representation in face recognition, target tracking and so on [7,16], this paper describe the fragment by the sparse histogram representation.

In the target area and the extended background area, n pieces of fragment images are randomly sampled to form the positive and the negative samples based on the positive and negative dictionary respectively. The samples are finally represented by the sparse representation. Then the positive samples as example as follows:

$$y = D_p a_p + e_p \qquad (1)$$

Wherein, D_p is the positive dictionary ; a_p is the positive sparse coefficient; e_p is the residual coefficient based on the positive dictionary.

Because the histogram representation not only can reflect the global statistical information of the coefficients, but also reflects the intuition of the coefficient distribution. So, In this paper, based on the positive and negative samples dictionary, the sparse histogram of the positive samples is formed respectively. In the same way, the sparse histogram of the negative samples is formed respectively.The difference of the object-background is described by the histogram, as follows:

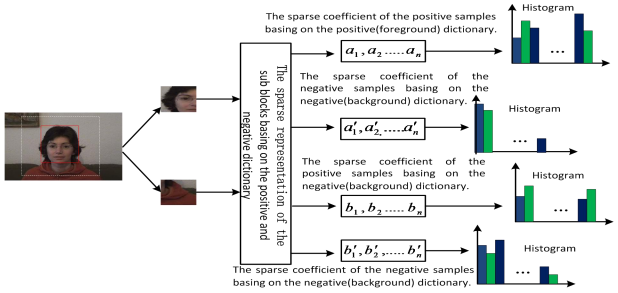

Figure 1: The sparse histogram representation of the positive samples and the negative samples

B. The similarity measurement of discriminative analysis

In the similarity measurement, the representation of the sample is the key factor. the histogram intersection is used for the histogram measurement [17]. In tracking process, the similarity measurement is not accurate due to image quality, violent movement, occlusion or background interference and so on, which leads the object to deviate from the tracking frame and to tail after the background or other interfering objects. In this paper, the discriminant analysis of similarity measurement is introduced by the background.

Consequently, the sparse model of the target is represented based on the positive dictionary in the reference frame(F_0), the sparse model of the background is represented based on the negative dictionary in the reference frame(B_0), the sparse model of the object is represented based on the positive dictionary in current frame(F_1), and the sparse model of the background is represented based on the negative dictionary in current frame(B_1) . In this paper, beside the similarity between the F_0 and F_1 is considered, and the discrimination between the F_1 and the B_1 , the discrimination between the F_1 and the B_0 , the discrimination between the F_0 and the B_1 are also considered. Here, the matching rules is illustrated by the global model, as shown as in figure 2:

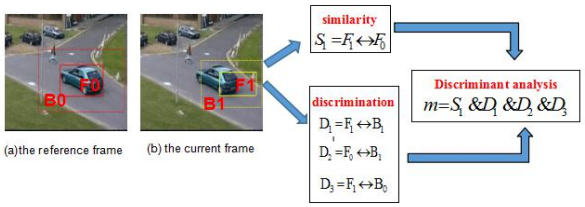

Figure2: The similarity measure based on discriminant analysis

Most of the present methods mainly consider the S1, in this paper, theD1, D2, D3 are also considered at the same time. Being different from the S1 which require high similarity, D1,D2,D3 require low similarity, that is the similarity of the D1, D2, D3 are lower, the accuracy of the similarity are better. The D1 is the discrimination of the target and the background. The D2 is the discrimination of the target in the current frame and the background in reference frame based on the positive dictionary. Usually when the target tracks to the background in current frame, because the background of the current frame and the background of the reference frame change not much, the similarity of the D2 is high, and in certain extent, it can prove the tracking error in current frame. The D3 is the discrimination of the target in reference frame and the background in current frame. If the similarity is high, in certain extent, the background is considered as the target, that is the target tracking is error.

By the analysis above, S_1 represents the similarity of the object-object, and D1,D2,D3 reflects the discrimination of the object and background, which represent the similarity from the opposite side. So in this paper, the similarity and

the discrimination are fused by the inverse operation. On the other hand, D1,D2,D3 are the different aspects of the discrimination, so they are fused by the product operation. The local similarity is formed by the fusing of the similarity and the discrimination, as follows as :

$$m = \frac{S_1}{D_1 D_2 D_3} \qquad (2)$$

Wherein, S_1 represents the similarity of the object between the reference frame and the current frame; D_1 represents the discrimination between the object template of the current frame and the background template of the reference frame; D_2 represents the discrimination between the object template of the reference frame and the background template of the current frame; D_3 represents the discrimination between the target template of the current frame and the background template of the current frame.

C. Verification of The Similarity

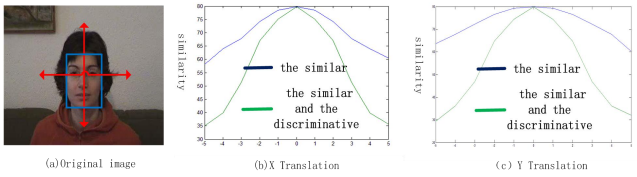

Figure 3. The comparison of the similarity

Figure3(a) is the original image, where the blue rectangle is the object template (initial position), the red cross is the center of the target frame, and the trajectories are moved along the x and y axes. “Fig 3 (b),(c)” is the variation trend of the similarity under the condition that the object rectangle moves along the x axis, y axis. With the shift change, the distance between the object and their template is further, the similarity between the object and their template is lower. So the Fig 3 illustrates that the difference of the similarity is larger, the feature that represent the object is more robust.

In Fig. 3 (a), the color of the face is different to the color of the hair, but the color of the eyebrows and eyes are similar to the color of the hair. So, the color of the object is different from the background and there is little background information. Because of the obvious differences between the object and the background, when the position of the object frame changes, the proportion of each color in the frame varies with the template. The traditional similarity measurement is not accurate because of the background in the target frame. When the object frame moves toward the surrounding, the background information in the target will gradually increase, which lead to increase the similarity of the object and the background. So the similarity of the similarity and the discrimination will become more accurate. As can be seen in the figure3, the similarity of the “similarity and discrimination fusion” is more accurate than the traditional measurement according to the changing of the curve in figure 3.

3. LEARNING DISCRIMINATIVE WEIGHTS

In tracking process, due to the influence of the background and noise, the contribution of the different local area in the similarity measurement is different, that is, some fragments are different from the background. Such as the 6 fragments in figure 4; on the contrary, the fragments are similar from the background, such as the 1 fragment.

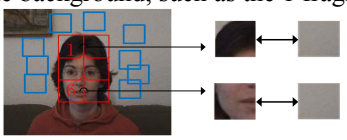

Figure 4：the discrimination of the fragment

In the similarity measurement, the fragment which is close to the background shows interference to the measurement; and the fragment which is different for the background is usually important, so the discrimination of the fragment to the background determines the contribution of the fragment in the similarity measurement. In this paper, to represent the discrimination of the fragment better, the single discriminative weight w_{dis} is proposed by the discrimination between the object and the background.

In the sparse representation of the sample, the residual coefficient e indicates the fitting degree of the sample to the dictionary, i. e., the closer the fit is, the smaller the e is, and vice versa. Therefore, in this paper, the discrimination of the fragment is determined by the residual coefficient e , as shown as follows:

$$w_{dis} = e_p / e_q \qquad (3)$$

Wherein, e_p is the residual coefficient of the fragment based on the positive dictionary; e_q is the residual coefficient of the fragment based on the negative dictionary.

4. TARGET TRACKING FRAMEWORK BASED ON DISCRIMINATIVE ANALYSIS AND LOCAL SPARSE REPRESENTATION

In this paper, the particle filter is used as a tracking framework. In this framework, the maximum posterior probability of the object state is the final object position, as follows:

$$x_t = \arg\max p(x_t \mid y_{1:t}) = \arg\max p(y_t \mid x_t^i) p(x_t^i \mid x_{t-1}) \qquad (4)$$

Wherein, x_t^i is the i sample in the t frame; $p(x_t \mid x_{t-1})$ is the state transition probability and is formed with the Gaussian distribution by the random throwing strategy.

A. Observation Model

The object is represented by the multi fragments and the local similarity measurement of each fragment respectively represent the similarity degree between the sample and the template. This paper fuses the local discriminative similarity of the each fragment to form the global apparent model, The similarity measure likelihood function of the i sample is as follows:

$$p(x \mid y) = \sum_{i=1}^{n} w_{dis} \left(\frac{S_1}{D_1 D_2 D_3} \right) \qquad (5)$$

Wherein, P is the posterior probability model; x, y is the state vector and the observation vector; w_{dis} is the

discriminative weight in the current frame; S_1 is the similarity measurement of the fragment; D_1, D_2, D_3 are the discirminative measurement of the fragment respectively.

B. Update of the fragment template

Obtain the optimal target state in the current frame, and to update the target template by the corresponding image fragment .

$$\begin{cases} t_i = kt_i + (1-k)y_i, & if \quad s(t_i, y_i) > \eta \\ t_i = t_i & otherwise \end{cases} \quad (6)$$

Wherein, s is the Bhattacharyya distance, i.e, $s(a,b) = \sum_{x \in X} \sqrt{a(x)b(x)}$; η =0.75 is the Empirical thresholds; k=0.8 is the updating coefficient.

5. EXPERIMENTS

In order to evaluate the robustness of the tracking algorithm in complex scenes, the seven video sequences were tested. These video sequences consist of some changes: the local occlusion, the illumination transformation, the attitude transformation and the complex background. In order to illustrate the effectiveness of the proposed algorithm better, this paper compares the five existing tracking algorithms, including Frag[12], MIL[4], WLCS[14], DFT[6], and LOT[13]. For each video sequence, the target position is manually marked at the initial frame. Our regularization parameter is that the histograms were computed using 256 bins and N=400 particles are used for test.

The time of the proposed method is also evaluated and compared with that of other methods. The results are listed in Table 1. The LOT method is the fastest, and it only need 0.18s to deal with one frame. Our method and the frag is middle in all methods.

Table 1. Average tracking time of different methods

Methods	Frag	WLCS	DFT	LOT	MIL	Our
Time(second/frames)	0.45	0.48	0.56	0.18	0.51	0.41

A. Quantitative Experiment

In this paper, the Euclidean distance calculation is used to calculate the center distance error between the target and the true position of the target to be the standard measurement. The smaller the value, the more accurate the method. Figure 6 demonstrate the center error plots 4 videos in all and the table 2 shows the average center error of the various algorithms over 7 videos. The tracking results on the challenging sequences show that the proposed algorithm has the minimum average center location error.

Table 2. Average center location error

	Frag	WLCS	DFT	LOT	MIL	Our
Caviar	5.415	2.607	18.299	2.2605	77.774	**1.674**
Women	24.196	9.569	43.501	32.385	31.584	**4.703**
DavidIndoor	71.158	93.477	61.349	59.292	60.985	**3.654**
DavidOutdoor	74.846	175.05	280.39	24.451	197.95	**5.757**
Men	**47.124**	69.947	160.83	**36.515**	117.45	**56.154**
Deer	101.69	10.046	412.21	72.974	211.93	**8.526**
Dollar	54.872	13.476	50.716	57.872	73.85	**3.472**
Highway	21.173	8.871	35,734	27.390	17.692	**7.615**

B. Qualitative experiment

The qualitative experiment of those six algorithms based on the four videos is shown, as follows as the figure 7.

The figure 7 shows that the proposed algorithm comparing with other methods is efficient in occlusion, scale change, background disturbance and abrupt motion.

6. CONCLUSION

To summary, the object tracking based on the local discriminative analysis is proposed under the framework of particle filters. Compared with traditional methods, this paper have many advantages as following: Firstly, based on the sparse representation of the fragments, the fragment measurements combined with the similarities and the discriminations is proposed by the discriminative presentation of the sparse coefficient. Secondly, based on the difference of the fragment in the video, the discriminative weight of fragment is proposed to improve the accuracies of the measurements. At the end, the similarity measurements of the samples are effectively calculated by fusing different fragment measurements. These experiments proved that the algorithm is more robust and more accurate than those popular algorithm in the challenging video. In further studies, we will describe the object by combining with the local and global based on the sparse representation, In addition, we will describe the target by fusing some different representation.

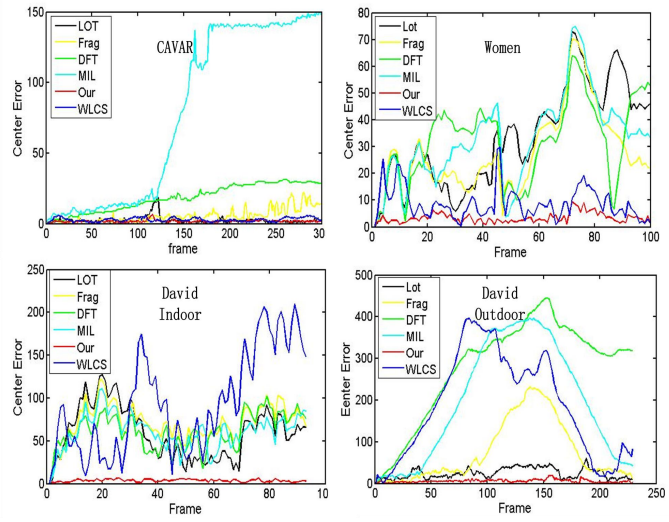

Figure 6: Center error plots of difference algorithm on 4challenging image sequence.

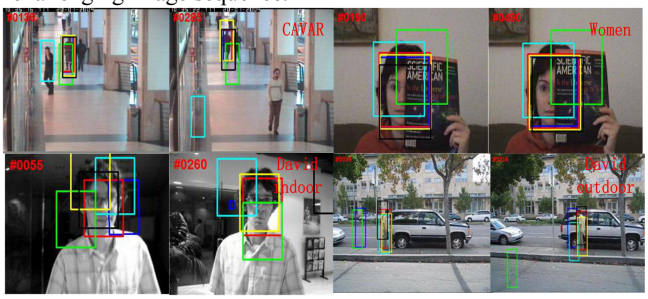

Figure 7: Tracking results on seven challenging sequence. RFT[19] (Green), DFT[22] (Blue), LOT[23] (Magenta),LSS[24] (Yellow), RCT[25] (Cyan) and our algorithm (Red).

REFERENCES

[1] L. Hongliang, Z. Shenghua, and L. Hongwei, et al. "An integrated target detection and tacking algorithm with constant track false alarm rate," *Journal of Electronics &Information Technology*, 2016, 38(5): 1072-1078. doi: 10.11999/JEIT150638.

[2] Z. Markus, N. Thomas, and K. Andreas, "Tracking human locomotion by relative positional feet tracking". *IEEE Virtual Reality (VR)*, USA, 2015, 317 - 318.

[3] P. Peng, B. Erik, and B. Hai, "Encoding color information for visual tracking:algorithms and benchmark". *IEEE Transactions on Image Processing*, 2015, 24(12). 5630-5644. doi:10.1109/TIP.2015.2482905.

[4] B. Babenko, M. H. Yang, S. Belong, "Robust object tracking with online multiple instance learning". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8), pp. 1619 – 1632.

[5] S. L. Laura and L. M. Erik, "Distribution fields for tracking," I*EEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2012, 1910-1917.

[6] C. Bao, Y. Wu , H. Ling, and H. Ji, "Real time robust L1 tracker using accelerated proximal gradient approach". *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, USA, 2012, 1830-1837.

[7] X. Mei, H. Ling, and Y. Wu, "Efficient minimum error bounded particle resampling LI tracker with occlusion detection," *IEEE Transactions on Image Processing*, 2013, 22(7):2661-2675.

[8] J. Kwon, and K. M. Lee, "Highly non-rigid object tracking via patch-based dynamic appearance modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(10):2427-2441.

[9] B. Liu, J. Huang, and L. Yang , et al, "Robust tracking using local sparse appearance model and k-selectio]," *IEEE Conference on Computer Vision and Pattern Recognition.* Colorado Springs,CO, USA,2011, 201:1313-1320.

[10] X. JIA, H. Lu, and M. H. YANG, "Visual tracking via adaptive structural local sparse appearance mode," *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, USA, 2012, 1822-1829.

[11] A. Adam, E. Rivlini , and I. Shimshoni, "Robust fragment-based tracking using the integral histogram," *IEEE Conference Computer Vision and Pattern Recognition (CVPR),* USA, 2006, 798-805.

[12] S. He, Q. Yang, and M. H. Yang, "Visual tracking via locality sensitive histograms," *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, USA, 2013, pp. 2427 – 2434.

[13] S. Oronr , and B. H. Aharon, "Locally orderless tracking," *International Journal of Computer Vision*, 2015, 111(2), pp:213-228.

[14] D. Wang, H. C. Lu, and C. J. Bo, "Visual tracking via weighted local cosine similarity," *IEEE Transaction on Cybernetics,* 2015, 45(9).

[15] P. M. Qaguiar, and J. M. Moura, "Figure-ground segmentation from occlusion," *IEEE Transactions on Image Processing*, 2005,14(8).

[16] B. H. Zhuang, H. C. Lu , X. Z. YiIan, and W. Dong, "Visual tracking via discriminative sparse similarity map," *IEEE Transaction on Image Processing,* 2014, 23(4).

[17] D. Y. Bi, T. Ku, and Y. F. Zha, et al, "Scale-adaptive object tracking based on color names histogram," *Journal of Electronics & Information Technology,* 2016, 38 (5): 1009-1106. doi: 10.11999/JEIT150921.