Target Tracking Algorithm Based on HOG Feature and Sparse Representation

Ming Li[a] and Qingsong Fang[b]

School of Computer and Communication, LanZhou University of Technology, Lanzhou 730050,China

[a]lim3076@163.com, [b]fqs_1991@yeah.net

Keywords: visual tracking, HOG feature, sparse representation, classifier

Abstract. In this paper, we propose a novel algorithm to deal with the problem of visual tracking in some challenging situations, which is based on HOG feature and sparse representation. First of all, describe target according to the HOG feature; secondly, construct the appearance model of target with the sparse representation, and then predict the target position on the basis of the particle filter method. At last, apply Naive Bayes classifier to track target. The experiment results show that the proposed algorithm is superior in accuracy than the classical tracking algorithm and has better robustness in the scene that contains the target posture changes, illumination variations and occlusion.

Introduction

Target tracking [1] has been concerned by many scholars as one of the most important research subjects in the field of Computer Vision [2]. In recent years, sparse representation has been gradually applied in target tracking, which is inspired by the human face recognition method based on sparse representation.

Sparse representation is a method with the minimum criterion l_1 as the core and based on template matching. Mei et al. propose a target tracking method (l_1 tracker) based on the minimization of the l_1 paradigm in the literature [3]. This tracking algorithm has the disadvantages of high computational complexity and it is easy to appear drift. The kind of tracking method based on global information can only be used to obtain the global features of the image [4-5], it is more suitable for a complete target appearance, but is difficult to solve problems of local image when it changes rapidly. Another tracking method based on local information only can obtain local feature of image well [6-7]. These local features can well apply to target with some changes, but are susceptible to suffer from the impact on posture changes and illumination changes. Therefore, its stability is not high.

In this paper, we propose a target tracking algorithm based on HOG feature and sparse representation, which combine the characteristics of global information and local information under the framework of particle filtering. A target template dictionary constructed by sparse representation can clearly distinguish target samples from non target samples.

Theoretical principles

Particle filter based on the basic idea of the optimal Bayesian estimation is a recursive algorithm for solving the posterior probability. Typically, in the framework of particle filter, the moving state of the target is expressed by the Gauss function, namely that it is Gauss function to determine the particle selection.

Native Bayes classifier [8] is a fast, simple and high efficiency classifier. In the Naive Bayesian classifier we assume that each attribute of a given sample is independent from each other, namely that there is no dependency between the attributes under the premise of knowing the variable. Naive Bayes classifier has characteristics of small classification error, simplicity, efficiency and stability, thus the classifier has a good application value on the research.

The tracking image has only two properties during the process of tracking, which are target and background. Assume that the prior probability of target and background is $p(y)$, where $p(y=1)=p(y=0)$, $y \in \{0,1\}$ which represents the type of sample, is a binary variable. $y=1$ represents the positive samples, $y=0$ represents the negative samples. Binary Naive Bayesian classifier is shown below:

$$H(X)=\log(\frac{\prod_{i=1}^{N} p(x_i \mid y=1)\, p(y=1)}{\prod_{i=1}^{N} p(x_i \mid y=0)\, p(y=0)})=\sum_{i=1}^{N}\log(\frac{p(x_i \mid y=1)}{p(x_i \mid y=0)}) \tag{1}$$

The confidence value of candidate sample in the classification is that:

$$l=\sum_{i=1}^{N}\log\left\{\frac{\frac{1}{\sigma_i^1}\exp(\frac{(x_i-\mu_i^1)^2}{2(\sigma_i^1)^2})}{\frac{1}{\sigma_i^0}\exp(\frac{(x_i-\mu_i^0)^2}{2(\sigma_i^0)^2})}\right\} \tag{2}$$

Where μ_i^1 is the mathematical expectation of positive training sample, σ_i^1 is the variance of positive training sample. While μ_i^0 is the mathematical expectation of negative training sample, σ_i^0 is the variance of negative training samples.

Our visual target tracking algorithm

Use HOG feature to describe target. The core idea of HOG feature is that use a gradient or the edge orientation density distribution to describe the appearance and shape of the local target in an image effectively, combine representation information and edge information of target together, thereby improve the robust of the tracking system. The detailed calculation procedure is shown as follows:

1) Use the unified center template [-1, 0, 1] to calculate the gradient $p_x(x,y)$ and $p_y(x,y)$ which are in the direction of horizontal and vertical respectively.
2) Calculate the size and direction of each pixel module according to the below formulas:

$$Norm=\sqrt{p_x^2(x,y)+p_y^2(x,y)} \tag{3}$$

$$or\,i\,ent=\arctan(\frac{p_y(x,y)}{p_x(x,y)}) \tag{4}$$

3) Divide the image into smaller regions at equal size, named cells, and then forms these cells to a block region and normalize its gradient value to generate the final image feature.

Construct the sparse representation model for target. Refer to the method in article [9-10] to construct the feature vectors of target appearance. Assume that image sequence is a n frames training image $T=[T_1,T_2,...,T_n]$, regard it as the target template set. Where $T_i \in R^{d\times d}$. Adopt a set of local image blocks to compute the local dictionary and then use the local dictionary to encode the local image blocks which exist in the target candidate region. Combine all target templates to obtain the final sparse coding dictionary:

$$D=[d_1,d_2,...,d_{(n\times N)}] \in R^{d\times(n\times N)} \tag{5}$$

Where $d=\omega\times\omega$ denotes the dimension of image block vector, n indicates the number of templates, N is the number of local image blocks obtained by sampling in each template. Since the local dictionary of each target template is obtained by vectoring and L_2 norm-normalize all the local image blocks t, and each local block represents one fixed part of a target object. Therefore, combine all the local block can fully represent the complete structure of target. Extract the partial image block of the sample which acts as a candidate target and belongs to $y \in R^{d\times d}$, and make use of the above method to

calculate the corresponding local dictionary, thus get the sampling representation of candidate target samples:

$$Y = [y_1, y_2, ..., y_N] \in R^{d \times N} \tag{6}$$

According to the sparse representation method, the local image block of the sample can be represented by a sparse encoding dictionary D. We can get the sparse coding of each image block in the dictionary D by solving the convex optimization problem:

$$\min_{b_i} \left\| y_i - Db_i \right\|_2^2 + \lambda \left\| b_i \right\|_1, s.t. \ \ b_i \geq 0 \tag{7}$$

Where y_i represents the $i-th$ vectorization partial image block, $b_i \in R^{(n \times N) \times 1}$ is sparse coding corresponded by partial image block. $b_i \geq 0$ means that all the elements of b_i are non-negative. The sparse coding of candidate sample can get as follow:

$$B = [b_1, b_2, ..., b_N] \in R^{(N \times n) \times N} \tag{8}$$

Make every target template in sparse dictionary vote for the partial image block of candidate sample, namely, weight the sparse coefficient of partial image block to obtain a weight vector $v_i[\upsilon_i^1, ..., \upsilon_i^N]$, where $\upsilon_i^j = \frac{1}{C} \sum_{k=1}^{n} c_i^{kj}$ is the voting weight that the $j-th$ image block in all target template vote for the $i-th$ image block of candidate target, C is the normalized parameter. Therefore, we can get the vote weighting matrix V of target template and candidate sample, where $V = [v_1, ..., v_M] \in R^{N \times N}$. Select $X = diag(V)$ as feature vectors, where $X \in R^N$. The reason lies in that the diagonal of weight matrix V represents the votes between the target template image block and the corresponding candidate sample image blocks.

Update the template. The template can only keep up with the objective observation model for a very short period of time during the tracking and will not adapt to complex scenes such as illumination variations or pose changes if it does not update on time. But if the update is too frequent, each update will introduce a small error which will result in drift problems with the accumulation of errors. In this paper, we use similar method with the article [10], to introduce the subspace learning method into the sparse representation method.

Update the classifier and collect samples. Fixed classifier cannot capture the appearance changes of target and background for a long period of time. Therefore, use dynamic update mechanism for updating the classifier in this article. Suppose that we collect the positive and negative samples in the tracking process to update the classifier after determining the target position at t time. Set a dynamic update parameter θ, and then retrain the classifier every θ frames.

Assume that the target location is l_t^*, the region that positive samples meet is $\left\| s^1 - l_t^* \right\| < r_1$. Collect positive samples in a circular area with a radius of 2 and a center of l_t^*. The region that negative samples meet is $r_1 < \left\| s^0 - l_t^* \right\| < r_2$, it means that collect negative samples in a circular ring with radiuses of r_1 and r_2. Where r_1 and r_2 can be set in the experiment.

The proposed algorithm. In this paper, combine HOG feature and particle filter as a method based on local information to mainly solve the occlusion problem of the target partially change. The combination of sparse representation and classifier, which can be used to process the problem of tracking macroscopically, is a method based on global information. Here are the detailed steps of the proposed algorithm:

Table 1.The proposed tracking algorithm

Input: m frame video image $\{f_t\}(t=1,2,\ldots,m)$.

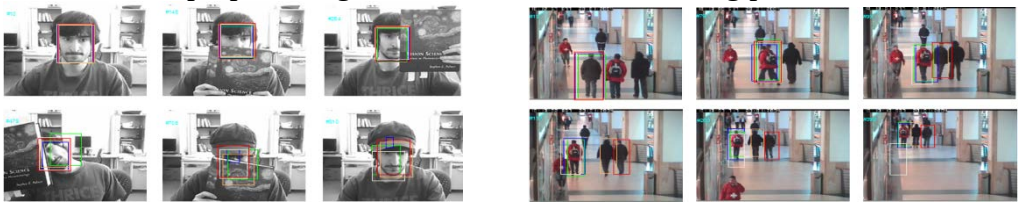
Output: the target position $\{x_t\}$ $(t=1,2,\ldots,m)$ obtained by each frame tracking.

1. Set up the initial m frames of image sequences as training images.
2. Calculate the HOG feature value of these m frames images separately, and then initialize the observation model of target.
3. Construct sparse coding dictionary D, and collect N_0 particles near target center.
4. Manually determine the state of the starting frame, use the particle filter estimates a set of locations which form the target of next frame, it is $l(x)=\{l_1(x),l_2(x),...,l_n(x)\}$, n is the limited value. Get a group of picture blocks according to the target position: $X^s=\{X_i^s, i=1,...,n\}$.
5. For each image block $\{X_i^s\}$, get their training samples $\{x_i, y_i\}$, Where $x_i \in R^N$ is a partial sparse feature vector, $y_i \in \{0,1\}$ is the tag of sample. Train the initial classifier to obtain the classification parameters $\{\mu_i^1, \sigma_i^1, \mu_i^0, \sigma_i^0\}_{i=1}^N$.
6. Calculate the target confidence value l_i of feature vector $\{x_i\}$, use the largest posterior probability to estimate the target state $X_i^* = \max_{x_i} l_i$, select the largest X_i^* from the image block collection $X^s=\{X_i^s, i=0,...,n\}$ to determine the final target state, thus obtain an accurate target position.
7. Re-sample the particle collection according to its weight size, and generate N_0 particles of the next frame tracking. Update the particles weights w_t^i, and normalize it according to $\sum_{i=1}^{N_0} w_t^i = 1$.
8. Update the target template, then determine whether we need to update the classifier according to the value of θ or not.

Experimental Results and Analysis

To evaluate the performance of the proposed algorithm, some standard video sets are tested. Experimental parameters are set as follows: the number of training images is $m=5$, the observed image that tracking algorithm used is a window whose size is 32×32, the number of samples is $N=30$, dynamic update parameter of the classifier is $\theta=5$, the number of particle is $N_0=500$, $r_1=30$, $r_2=50$. In order to further evaluate the proposed tracking method, compare the proposed method with four current mainstream algorithms: IVT [6], PCA [9], L1 and MIL [10]. IVT is blue, L1 is green, MIL is gray, PCA is orange, our tracker is red. The experimental results are analyzed and illustrated from two aspects of qualitative and quantitative.

Qualitative Analysis. In the Faceocc2 video sequences, target go though the book, where the occlusion area reaches to 70% of the face visual areas. The experimental results are shown in Figure 1, image frames respectively are #10, #145, #264, #479, #706, and #810. After #479, the IVT method has an obvious deviation and the squeezing phenomenon. L1, MIL and PCA show small amplitude drift at #706, while the proposed algorithm shows a better tracking performance.

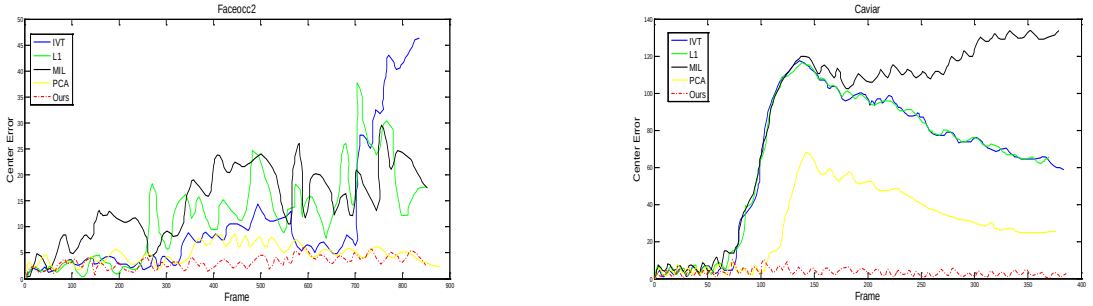


(a) (b)

Fig. 1 Tracking results of the sequence (a) Faceocc2 (b) Caviar

In the Caviar video sequences, the target undergoes two kinds of occlusions interference: similarity occlusion and non similarity occlusion. The tracking image frames respectively are #13, #79, #93, #137, #260 and #392. The tracking results are shown in Figure 1. IVT, L1 and MIL algorithms have been followed up to turn wrong when there are non similarity occlusions appearing at #79 in the first time. There is tracking error appearing in the PCA algorithm after the target at #137 experiences the occlusion of similar background. Obvious drift appear in MIL algorithm after #260. Because the proposed algorithm has the characteristics of local information, it is more suitable to process the occlusion condition. Therefore, the algorithm has an excellent tracking performance under occlusion condition.

Quantitative Analysis. In order to further analyze the performance of this algorithm, we take the center position deviation to quantitatively evaluate the performance of the algorithm. Center position deviation, which is used to quantitatively describe the tracker in this article and reference the performance of tracker, indicates the central location of the track results and the relative position deviation of test video sequences. Analysis results are shown in Table 2 and Fig.2. Table 2 shows the test results of arithmetic mean center error that 5 tracking methods generate in two test videos, traditional bold underlined line represents the best tracking results. Fig.2 is the line graph constructed by the center position deviation and frames.



(a) (b)

Fig. 2 Tracking center position error plot of the sequence (a) Faceocc2 (b) Caviar

Table 2 The numerical analysis of mean center error of the algorithms

	IVT	L1	MIL	PCA	Ours
Faceocc2	10.52	11.43	14.57	4.34	**3.94**
Caviar	65.75	66.08	84.15	37.64	**4.86**

Results Analysis. Learned from the experimental results, we know that the proposed algorithm can accurately track the target in Faceocc2 and Caviar video sequences. In the Faceocc2 video sequence, only the proposed tracker and PCA tracker maintain a stable tracking performance. The video sequences in which the parts of man wearing hat and face occluded by books keep changing. Because the algorithm uses the HOG feature to describe the target, the particle filter can accurately predict the next frame position of the target. In the Caviar sequences, the proposed tracker does not have tracking error. Compared with the other four trackers, the proposed tracker obtains the best tracking effect in Faceocc2 and Caviar video sequences. Taking into account overall performance, our tracker has the best effects in tracking process.

Summary

A robust tracking algorithm based on the combination of HOG feature and sparse representation, is the characterized integration of global information and local information method. In this paper, use the particle filter combined with HOG feature to predict the positions of target and apply the classifier combined with sparse representation to calculate confidence value for the position of each target, thus we can ultimately determine the exact position of target. HOG feature combined with particle filter is a good solution to the problem of occlusion in target tracking. Experiment results show that compared

with the current popular varieties of tracking algorithms, our algorithm has better robustness and accuracy on tracking problem. However, the integration of the global information and local information in this algorithm lead to slightly higher complexity of the algorithm. The focus of future research work is how to enhance the speed and reduce the computational in this algorithm.

Acknowledgements

This research is supported by the National Natural ScieceFoundation of China (No.61263019).

References

[1] H.L. Zhang, Sh.Q. Hu, G.Sh. Y, Video Object Tracking Based on Appearance Models Learning, Journal of Computer Research & Development. 52 (2015) 177-190.

[2] W. Lu, Zh.Y. Xiang, H.B. Yu, et al., Object compressive tracking based on adaptive multi-feature appearance model, Journal of Zhejiang University: engineeringence. 48 (2014) 2133-2138.

[3] X. Mei, H. Ling, Robust visual tracking using ℓ 1 minimization, Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2009) 1436-1443.

[4] D.W. Yang, Y. Cong, Y.D. Tang, Object Tracking Method Based on Particle Filter and Sparse Representation, Pattern Recognition & Artificial Intelligence. 26 (2013) 680-687.

[5] T. Bai, Y.F. Li, Z. Shao, Online visual object tracking with supervised sparse representation and learning[C]// 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV). 2014.

[6] J.H. Xiang, H. Fan, J. Xu, et al., Object tracking based on local sparse representation[J]. Journal of Huazhong University of Science & Technology. 2014.

[7] T. Bai, Y. Li, Robust Visual Tracking Using Flexible Structured Sparse Representation, IEEE Transactions on Industrial Informatics. 10 (2014) 538-547.

[8] S. Taheri, M. Mammadov, Learning the naive bayes classifier with optimization models, International Journal of Applied Mathematics and Computer Science. 23 (2013) 787-795.

[9] P.Y. Dai, J.X. Hong, C.H. Li, X. J. Zhan, A discriminant target tracking algorithm based on sparse representation, Journal of Xiamen University. 53 (2014) 477-483.

[10] X. Jia, H. Lu, M.H. Yang, Visual tracking via adaptive structural local sparse appearance model[C]// IEEE Conference on Computer Vision & Pattern Recognition. (2012) 1822-1829.