

Random Ferns with HOG-LBP Feature and Online Multi-instance Learning for Complex Tracking

JI Pei-pei, CHEN Ken^{*}, LIU Zhe, WU Ying

(Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: In this paper, a target modeling approach using random ferns with HOG-LBP feature and online multi-instance learning is presented. The proposed HOG-LBP feature vector with a target image block is utilized to initialize and later form new random ferns which ultimately serve as the target detector and tracker. The target model updating is attained through online multi-instance learning with a fern pool. The proposed approach is put to test on the given number of standard test sequences, and compared with other recently better known tracking algorithms, manifesting its more satisfying performance and objective strength in tracking in complex scenes.

Key words: visual tracking; target modeling; random ferns; HOG-LBP feature; multi-instance learning

CLC number: TN919.8

Document code: A

Article ID: 1001-5132 (2015) 04-0042-06

Detection based visual tracking with sample learning focuses on the building of self-adaptive target appearance model. Normally, in the initial tracking period there is not a sufficient number of samples for learning, but more and more samples are accumulated as the tracking proceeds. Considering these conditions, some researchers have put forward online learning based tracking algorithms. Saffari et al^[1] proposed the online random forest classification model. Ozuysal and Villamizar et al demonstrated that the random ferns approach can achieve classification performance at much lower cost because this approach is much more time-saving in computing the branch at each node due to using LBP features^[2-4]. Grabner et al proposed an online learning method called “Online Boosting” using the adaboost learning approach in which the new samples update

the feature pool of a classifier in a batch-processing manner^[5]. Later, they further brought forward SemiBoost tracking-and-learning framework by combining the semi-supervising learning and adaboost learning^[6-7]. Some researchers have made efforts to combine the online boosting learning framework with multiple instance (MIL) sampling method, one of which is the tracking algorithm based on the MILBoost online learning scheme proposed by Babenko^[8-9]. Another approach is proposed by Hong et al in which an offline tracking is implemented using tree-structured graphical model identification ensuring better tracking robustness in large motion, shot changes, occlusion, etc^[10]. As one of the most popular learning method, the multiple instance learning (MIL) has seen its most recent applications in group activities recognition^[11].

Received date: 2014-10-31.

JOURNAL OF NINGBO UNIVERSITY (NSEE): <http://journalg.nbu.edu.cn/>

Foundation items: Supported by National Natural Science Foundation (2011ZX03002-004-02); Supported by Natural Science Foundation of Ningbo (2014A610065); Supported by Teaching and Researching Project of Ningbo University (XKXL1308).

The first author: JI Pei-pei (1989-), female, Luoyang Henan, M.S. candidate, research area: video tracking. E-mail: helloji_130@163.com

***Corresponding author:** CHEN Ken (1962-), male, Chongqing, associate professor, research area: image and video processing, multimedia communication, intelligent control. E-mail: chenken@nbu.edu.cn

In this paper, based on the MILBoost framework, a random fern tracking algorithm is proposed using the HOG-LBP feature vector from the HOG feature space^[12].

1 Random ferns

The random ferns approach is a classification method that uses Bayesian probability model to estimate the target features. The class set is expressed with c_i , $i=1, \dots, H$ (in tracking procedure, H takes 2, that is, targets are classified as one class, and non-targets are classified as another class). And the binary feature set is connoted as f_j , $j=1, \dots, N$, and is extracted from an image block and then classified. The role played by a classifier is to seek the class that has the maximal conditional probability of the image block matching the target of interest, as formulated as

$$\hat{c}_i = \arg \max P(c = c_i | f_1, f_2, \dots, f_N), \quad (1)$$

where c stands for random variable of a class, and the conditional probability in Eq.(1) can be obtained from the Bayesian formula:

$$P(c = c_i | f_1, f_2, \dots, f_N) = \frac{P(f_1, f_2, \dots, f_N | c = c_i)P(c = c_i)}{P(f_1, f_2, \dots, f_N)}, \quad (2)$$

suppose that the a priori uniform distribution $P(C)$ exists, then the denominator in Eq.(2) is irrelevant to the class, the problem can thus be simplified as:

$$\hat{c}_i = \arg \max P(f_1, f_2, \dots, f_N | c = c_i). \quad (3)$$

Normally, the feature of the random ferns refers to the LBP feature that is obtained from comparing the image pixel values, which is prone to noise corruption. In this paper, a sample is obtained based on LBP feature from the HOG feature space, that is, a HOG-LBP feature f is proposed, and given as

$$f = \begin{cases} 1, & \text{if } \Omega(u, v, o) < \Omega(u^*, v^*, o^*), \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where Ω denotes the HOG feature space of an image block, u and v is the coordinates of the image plane, and o is the orientation angle of the gradient.

Considering the immensity of computing the combined probability distribution for all the binary features, all the features are divided into K groups according to the feature independent properties, the size of each group is defined as $S = N / K$, N is defined afore as the number of ferns, and the LBP feature in each group forms a fern. In doing so, the desired combined probability distribution can be expressed in Eq.(5):

$$P(f_1, f_2, \dots, f_N | c = c_i) = \prod_{i=1}^N P(F_k | c = c_i), \quad (5)$$

where $F_k = \{f_{\psi(k,1)}, f_{\psi(k,2)}, \dots, f_{\psi(k,S)}\}$, $k=1, \dots, K$, indicates the k -th fern, $\psi(k, j)$ is the random sequencing function. Fig.1 illustrates the fern constructing procedure. Each layer uses one feature to branch out. On a node, a conditional probability vector $P(F_k | c = c_i)$ can be obtained. Because the target model adopts the MILBoosting online learning framework^[8-9], the target model needs constructing a random fern pool from which K optimal ferns are selected and used for target detection purposes. Apparently, the total number of ferns in the random fern pool, M , is required to be much greater than K .

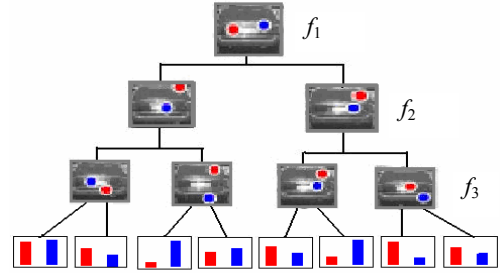


Fig.1 Illustration of fern construction

2 Tracking using HOG based multiple instance online random ferns

2.1 HOG-LBP feature

In recent years, the HOG filters alone have been applied in target detection^[13]. Fig.2 illustrates an image block from which some points are sampled and compared to form the HOG-LBP of a single fern. First, the HOG feature is extracted from the gray-scaled image

block, then randomly select two 3-D points, that is, (u, v, o) and (u^*, v^*, o^*) in the HOG feature space $\Omega(u, v, o)$ to compare their HOG value to obtain the LBP feature vector of the target. 8 gradient directions are evenly swept from a randomly selected point.

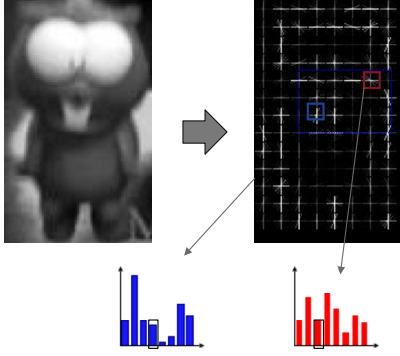


Fig.2 HOG-LBP feature extraction

2.2 Target model updating strategy

Prior to tracking, based on the initially labeled image block, the instance sampling method is used to select both the positive and negative samples. The detected on-target image block serves as a positive sample, and an off-target image block is selected as a negative sample. Each positive sample instance is turned into 20 positive sample instances by mixing and expanding noise which constitutes a positive sample pack. In the course of tracking, both the new positive and negative samples are first selected online continually, which are fed into the pool and used to update a single fern probability model, as given in Eq.(6). In the algorithm, the model updating factor, $\mu (0 \leq \mu \leq 1)$ as defined in Eq.(7), is used to control the rate of model updating, and Eq.(6) describes the a single random fern conditional probability updating process.

$$p^{(t+1)}(F_k | c) = \frac{\mu \text{Num}_{k,i,c}^{(t)} + \text{num}_{k,i,c}^{(t)}}{\mu T\text{Num}_{k,c}^{(t)} + T\text{num}_{k,c}^{(t)}}, \quad (6)$$

$$\mu = 1 - p'(c | X), \quad (7)$$

$$p'(c = 1 | X) = 1 - \prod_{x \in X} (1 - p'(c = 1 | x)), \quad (8)$$

$$p'(c = 0 | X) = \prod_{x \in X} p'(c = 1 | x), \quad (9)$$

where $p^{(t+1)}(F_k | c)$ means under condition of class c the probability of the feature of k -th fern F_k being i ; $\text{Num}_{k,i,c}^{(t)}$ computes the number of features being i in the

samples classified as c for k -th fern prior to t -th frame, while $T\text{Num}_{k,c}^{(t)}$ computes the total of $\text{Num}_{k,i,c}^{(t)}$ with respect to i . $\text{num}_{k,i,c}^{(t)}$ calculates the number of instances which have the feature being i in the new sample pack for the fern classified as c in t -th frame; $T\text{num}_{k,c}^{(t)}$ calculates the total number of $\text{num}_{k,i,c}^{(t)}$ with respect to i . In Eq.(7), the probability of sample pack belonging to class c , $p'(c | X)$, can be obtained using Eq.(8) and Eq.(9).

After updating each fern, next is to extract K optimal ferns in a batch-processing manner. Eq.(10) is used to seek the optimal fern in such a manner.

$$(h_k, \alpha_k) = \arg \max_{h \in H, \alpha} J(H_{k-1} + \alpha h), \quad (10)$$

$$h = \log \frac{P(F_k | C = 1)}{P(F_k | C = 0)}. \quad (11)$$

In this paper, the proposed algorithm divides the targets into two classes, that is, target class and non-target class. Therefore in Eq.(10), h can be obtained using Eq.(11), where α indicates the weighting of the random ferns. Upon completion of building the probability model, Eq.(12) is used as the operator for target detection in the next frame.

$$H(x) = \sum_{k=1}^K \alpha_k h_k(x). \quad (12)$$

As stated previously, K is the number of optimal ferns used to construct a target detection operator.

2.3 Strategy of target search and matching

Using the target model to track the target of interest in the designated area is the most critical step in the tracking algorithm. In the search window, first a number of locations are selected randomly, then using each location as the center, the rectangular image blocks with varying scales are formed as both the positive and negative samples. Next, prior to extracting the HOG-LBP feature, unify all the scales of image blocks to that of initial image block. Finally, the Sigmoid function $S(y) = 1 / (1 + e^{-y})$ is used to compute $H(x)$ for each sample and then map it to the range of 0 to 1^[8-9]. The mapping value is utilized as the conditional probability for approximately describing a single sample classification, which is expressed as below:

$$p(c|x) = S(H(x)). \quad (13)$$

Then a new and optimized target image block is obtained by comparing the probability of each sample in the sample pack.

$$i = \arg \max_{x_i \in X} p(c=1|x_i). \quad (14)$$

In Eq.(14), X denotes the sample pack collected in the course of target tracking, and i is the index of sample that has the maximal probability obtained in Eq.(13) which corresponds to the most probable target candidate.

3 Tests and analysis

Running on the platform of Macbook Pro with processor of 2.3 GHz Intel Core i7 and RAM of 8 GB, and under the C++ LLVM 5.0 and Matlab 2012b, the proposed algorithm is put to test on the given number of standard video sequences, including Car4, David, Lemming, Trellis. The algorithms running in C++ environment include OnlineBoostingTracker (OAB)^[5], SemiTracker (SBT)^[6], BeyondSemiTracker (BSBT)^[7],

MILTracker (MIL)^[8-9], and the algorithm running on Matlab and C++ combined is ComprehensiveTracker (CT)^[11]. In the proposed algorithm, the number of ferns in the fern pool, M , is 200, the number of the optimal ferns, K , is 50. The radius of the positive learning samples is 5 pixels, the number of positive samples is 45; the radius of the negative learning samples is 50 pixels excluding that of image block, the number of negative samples is 65. In the period of target searching, the number of the randomly selected locations is 15, and at each of which 5 samples with various scales are selected. The number of features for a single fern, S , is 8.

Fig.3 demonstrates the visual tracking effect on the four given sequences using the selected referenced algorithms mentioned above. In the figure, each row contains 5 representative frames taken from Car4, David, Lemming and Trellis, respectively. All the sequences contain illumination change and partial occlusion, and the proposed algorithm shows the better performance in all the four sequences. In Lemming and Trellis, the background scene is relatively more complex, making

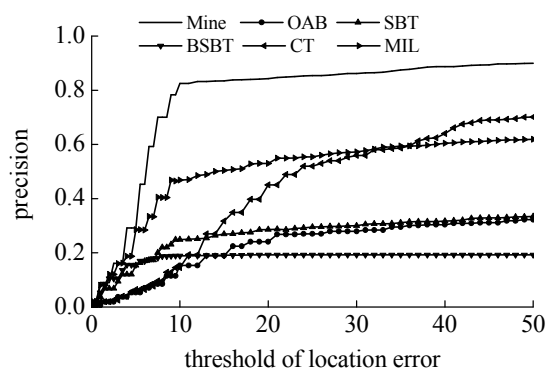


first row is from Car4, second row David third row lemming fourth row Trellis

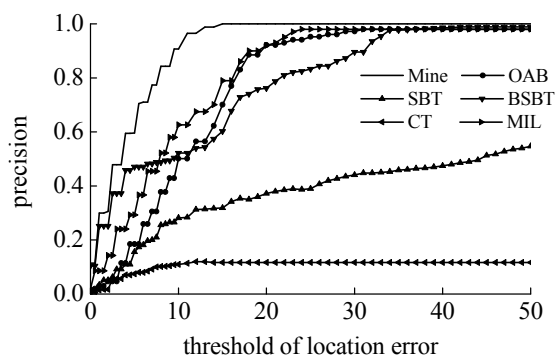
Fig.3 Tracking performance comparison

the proposed algorithm result with tracking deviation, suggesting that the proposed HOG-LBP feature based algorithm is not as strong as other tested approaches in terms of rotation-resistance. Finally, due to multi-scale sampling used, the proposed algorithm possesses certain compatibility in tracking the target with variant scale.

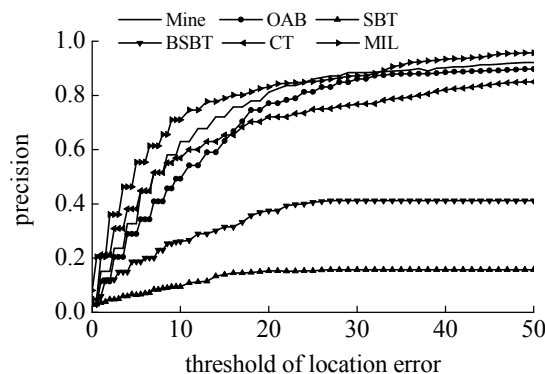
To further analyze the algorithmic performance, Fig.4 shows the tracking performance of the proposed algorithm (Mine) comparing with other 5 tracking algorithms which are Online Boosting Tracker (OAB)^[5], SemiTracker (SBT)^[6], beyond Semi Tracker (BSBT)^[7], CompressiveTracker (CT)^[14], and MILTracker (MIL)^[8-9]. In the figure, the abscissa indicates the tracking error threshold (in pixels), the ordinate marks the tracking precision (in percentile). it can be seen that the proposed approach performs better in Car4, David and Lemming. In contrast, in Trellis, owing to dim illumination, plus significant rotation occurrences, the proposed approach is not as strong as CT, but still better than the rest algorithms.



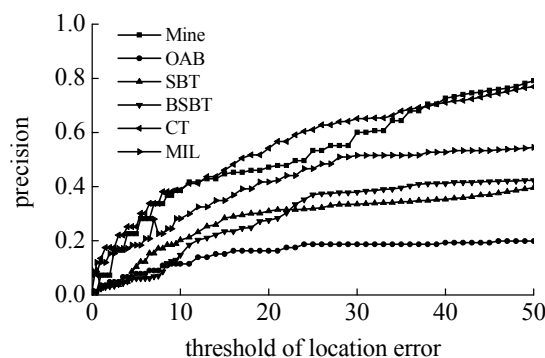
(a) Car4



(b) David



(c) Lemming



(d) Trellis

Fig.4 Tracking performance evaluation

4 Conclusions

In this paper, a detection based tracking approach is presented using the HOG-LBP target features and random fern classification model in combination with multi-instance Amir learning framework. The presented approach is put to test on the given number of standard video sequences against other more commonly used tracking algorithms in a variety of scene complexity factors. The latest proposed evaluation method is also applied for tracking performance comparison purposes. The results indicate that the proposed algorithm possesses better performance in general, while still has room in improvement in rotation resistance. The tests demonstrate the generally satisfying tracking robustness in complex scene circumstances.

References:

- [1] Saffari A, Leistner C, Santner J, et al. Online random

- forests[C]//Proceeding of the 12th International Conference on Computer Vision: Kyoto Japan, 2009: 1393-1400.
- [2] Ozuysal M, Calonder M, Lepetit V, et al. Fast keypoint recognition using random ferns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 448-461.
- [3] Ozuysal M, Fua P, Lepetit V. Fast keypoint recognition in ten lines of code[C]//IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis Minnesota USA, 2007:1-8.
- [4] Villamizar M, Moreno-Noguer F, Andrade-Cetto J, et al. Efficient rotation invariant object detection using boosted random ferns[C]//IEEE Conference on Computer Vision and Pattern Recognition: San Francisco USA, 2010: 1038-1045.
- [5] Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting[C]//IEEE Conference on Computer Vision and Pattern Recognition: New York USA, 2006: 260-267.
- [6] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking[C]//European Conference on Computer Vision: Marseille France, 2008: 234-247.
- [7] Stalder S, Grabner H, Gool L. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition[C]//International Conference on Computer Vision: Kyoto Japan, 2009:1409-1416.
- [8] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8):1619-1632.
- [9] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]//IEEE Conference on Computer Vision and Pattern Recognition: Miami, Florida USA, 2009:983-990.
- [10] Hong Seunghoon, Han Bohyung. Visual tracking by sampling tree-structured graphical models[C]//European Conference on Computer Vision, Florence Zurich Switzerland, 2014:1-16.
- [11] Antic B, Ommer B. Learning latent constituents recognition of group activities in video[C]//European Conference on Computer Vision, Florence Zurich Switzerland, 2014:33-47.
- [12] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, San Diego USA, 2005: 886-893.
- [13] Ahmed E, Shakhnarovich G, Maji S. Knowing a good HOG filter when you see it: Efficient selection of filters for detection[C]//European Conference on Computer Vision, Florence, Zurich, Switzerland, 2014:80-94.
- [14] Zhang Kaihua, Zhang Lei, Yang M H. Real-time Compressive Tracking[C]//European Conference on Computer Vision, Florence Italy, 2012:866-879.

结合 HOG-LBP 特征及多实例在线学习的随机蕨目标跟踪方法

吉培培, 陈 愚*, 刘 哲, 吴 盈

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 提出了通过联合随机蕨与 HOG-LBP 特征建立目标模型的方法。首先利用选定目标图像块中的 HOG-LBP 特征向量完成初始化及在其后的过程中产生新的随机蕨, 该随机蕨用于对兴趣目标的检测和跟踪。目标模型的更新通过多实例在线学习与更新蕨池实现。提出的方法在选定 6 个标准视频序列进行测试, 测试结果与现在较流行的 OnlineBoostingTracker、MILTracker 等跟踪算法进行了比较和分析。结果表明, 在各种复杂环境下, 本研究提出的方法具备良好的跟踪鲁棒性, 在抗局部遮挡和光照变化方面具有一定的优势; 同时算法具备一定的抗尺度变化能力; 在抗旋转方面, 该算法仍有一定的可改善空间。

关键词: 视觉跟踪; 目标模型; 随机蕨; HOG-LBP 特征; 多示例学习

(责任编辑 史小丽)