

Visual Tracking Based on Ensemble Learning with Logistic Regression

Xiaolin Tian^(✉), Sujie Zhao, and Licheng Jiao

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center of Intelligent Perception and Computation, International Collaboration Joint Lab in Intelligent Perception and Computation, Xidian University, Xian 710071, Shaanxi, China
xltian@mail.xidian.edu.cn

Abstract. In this paper, we propose a novel visual tracking method based on ensemble learning using logistic regression model. We adopt logistic regression to achieve ensemble classifier to deal with object tracking problem. By using fast computable features, our approach learns the appearance of the target during tracking. And thus, the proposed method is able to adapt online to target appearance changes and its surrounding background. Moreover, ensemble learning converts rough rules of thumb into highly accurate prediction rule. Experimental results show that our method outperforms relative trackers.

Keywords: Visual tracking · Logistic regression · Ensemble learning

1 Introduction

In computer vision field, visual tracking has been an important branch and has wide applications including video surveillance, robotics, autonomous navigation and human computer interaction [1]. Based on the discriminative model, the tracking problem can be treated as a classification task [2]. Hough-based tracking of non-rigid objects (HBT) [3] locates the support of the target through back projection from a Hough Forest. Multiple instance learning (MIL) [4] learns a discriminative classifier from positive and negative bags of samples. Struck [2] applies a structured output (support vector machine) SVM to directly predict the change in object location between frames, instead of using a labeler. Because of the strong convexity and probabilistic underpinnings, logistic regression (LR) is widely studied and used in many applications [5]. Compared with support vector machine, the advantages of LR are its posterior model for model selection and its probabilistic output for uncertainty prediction [5], which can be used for comparing classifier outputs. Different from the previously proposed methods, we introduce ensemble learning based on logistic regression model to deal with the visual tracking problem. The remaining part of this paper is organized as follows: Sect. 2 discusses the proposed method. Experiment results are described in Sect. 3, and Sect. 4 concludes this paper.

2 The Proposed Method

2.1 Logistic Regression Classifier

Let $x \in R^N$ denote a vector of explanatory or feature variables, and $y \in \{-1, +1\}$ denotes the associated binary output. Logistic regression attempts to find a separating hyperplane in feature space, parameterized by normal vector $w \in R^N$, which separates the two classes [6]. The posterior label probability is modeled as:

$$P(y|x, w) = \frac{1}{1 + \exp(-yx^T w)} \tag{1}$$

Suppose we are given a set of training or observed examples $x = \{x_1, x_2, \dots, x_M\}$ and their label $y = \{y_1, y_2, \dots, y_M\}$, the model parameter w can be found by maximum likelihood estimation from the observed examples. The maximum likelihood estimate minimizes the average loss [7]:

$$l_{avg}(w) = \frac{1}{M} \sum_{i=1}^M \log(1 + \exp(-y_i w^T x_i)) \tag{2}$$

In many cases, the maximum-likelihood estimator may overfit to the training data [6]. To reduce overfitting, penalized likelihood methods based on l_2 -regularization seek to minimize a version of:

$$J(w) = l_{avg}(w) + \lambda \|w\|_2^2 \tag{3}$$

where $\lambda > 0$ is the regularization parameter. There are many methods for training logistic regression models. In fact, most unconstrained optimization techniques can be considered [8]. Quasi Newton [9,10] is used to solve the weight W in our paper.

2.2 Weak Classifier

Haar-like feature is used in the proposed method. This feature is a simple rectangle features proposed by [11,12]. Each weak classifier h_k is composed of a haar-like feature f_k and four parameters $(\mu_+, \sigma_+, \mu_-, \sigma_-)$ that are estimated online [4]. The classifiers return the log odds ratio:

$$h_k(x) = \log\left[\frac{P(y = +1|f_k(x))}{P(y = -1|f_k(x))}\right] = \log\left[\frac{P(f_k(x)|y = +1)P(y = +1)}{P(f_k(x)|y = -1)P(y = -1)}\right] \tag{4}$$

where $P(f_k(x)|y = +1) \sim N(\mu_+, \sigma_+)$ and similarly for $y = -1$. We let $P(y = +1) = P(y = -1)$ and use Bayes rule to compute the above equation. When the weak classifier receives new data $\{(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)\}$, we use the following update rules:

$$\mu_+ \leftarrow \gamma \mu_+ + (1 - \gamma) \frac{1}{M} \sum_{i|y_i=+1} f_k(x_i) \quad (5)$$

$$\sigma_+ \leftarrow \gamma \sigma_+ + (1 - \gamma) \sqrt{\frac{1}{M} \sum_{i|y_i=+1} (f_k(x_i) - \mu_+)^2} \quad (6)$$

where $0 < \gamma < 1$ is a learning rate parameter. The update rules for μ_- and σ_- are similarly defined.

2.3 Ensemble Learning Based on Logistic Regression Framework

The proposed ensemble learning method uses logistic regression to optimize their weights of weak classifiers. Figure 1 shows the relevant steps. Ensemble learning refers to boosting the performance of a classifier by training many weak classifiers and combining them with weights [13]. When it is difficult to design a high performance classifier, boosting is particularly useful way for coping with the problem and providing simple decision rules to perform slightly better than random guessing. In general, the final strong classifier is a linear combination of the weak classifiers. The boosting algorithm is to find a way to boost a set of simple (weak) classifiers into a much stronger classifier through a certain learning method [13].

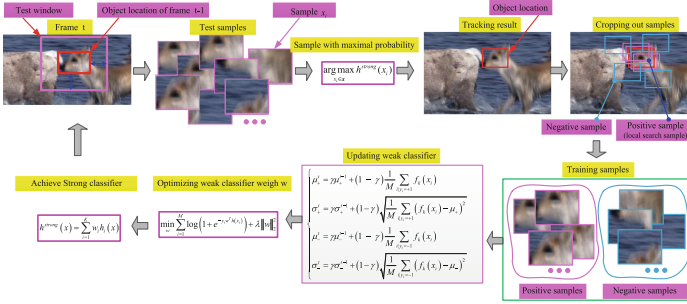


Fig. 1. Tracking model based on ensemble learning with logistic regression

Considering the simplicity and computational efficiency, we crop out a set of image patches within a test area based on the tracker location of previous frame when a new (current) frame arrives. The image patch with the highest posterior probability given by the boosting classifier is determined as object patch, and its location is defined as the object location. The prediction function in the algorithm is

$$h^{strong}(x) = \sum_{i=1}^K w_i h_i(x) = w^T h(x) \quad (7)$$

where $h_i(x)$, $i = 1, 2, \dots, K$ is the ‘better’ weak classifiers.

Based on the objection location, we can acquire the positive and negative samples by cropping out several image patches. Each image patch is viewed as the training sample and corresponds to a feature vector in our case. The weak classifier parameter is updated according to Eqs. 5 and 6. We select some better weak classifiers and provide an appropriate weight for each of them by logistic regression.

$$\min_w \sum_{i=1}^M \log(1 + \exp(-y_i w^T h(x_i))) + \lambda \|w\|_2^2 \quad (8)$$

Equation 8 reduces entirely the error between the predicted label and the true label. Accordingly, the weights of weak classifiers are determined.

3 Experiments

We empirically set $\gamma = 0.95$, $N = 250$ and $K = 100$ in our experiments. To evaluate the effectiveness of the proposed approach, we compare our tracker against state-of-the-art algorithms (CT [2], CXT [14], DF [15], MIL [4], SCM [16], Struck [2], TLD [17] and VTD [18]) on several publicly available challenging image sequences. They cover various challenging situations (partial occlusion, illumination variation, pose change, motion blur, etc.) for object tracking.

Table 1 reports the average center location errors (in pixels), where a smaller value indicates a more accurate tracking result. Table 2 reports overlap success rate (%) with a threshold of 0.5, where the larger average scores indicate more accurate results. The provided qualitative comparison on seven challenging sequences are shown in Fig. 2. It confirms that our tracer handles the following situations:

Table 1. Average center location errors (in pixels). The red fonts and the blue fonts indicate the best and the second best performances respectively.

Sequence	CT	CXT	DF	MIL	SCM	Struck	TLD	VTD	Ours
Basketball	89	215	18	92	53	118	269	6	10
David3	89	222	51	30	73	107	281	67	13
Football	12	13	9	12	17	17	14	14	12
Jogging	92	6	31	96	132	62	7	83	5
Liquor	186	132	221	142	99	91	100	60	57

Occlusions and Deformation: Occlusion is one of the crucial problems in visual tracking. Figure 2(a), (d) and (e) show the performance of all trackers when the tracking object suffers partial and heavy occlusions. Only CXT, TLD and our method can keep track of the target in the Jogging sequence. Our method even successfully deals with twice occlusion while other approaches

Table 2. Overlap success rate (%) with a threshold of 0.5. The red fonts and the blue fonts indicate the best and the second best performances respectively.

Sequence	CT	CXT	DF	MIL	SCM	Struck	TLD	VTD	Ours
Basketball	25.93	2.48	71.59	27.45	60.28	10.21	2.48	92.41	81.51
David3	34.92	13.89	74.21	68.25	48.02	33.73	10.32	48.41	84.52
Football	78.45	65.19	84.25	73.76	57.18	66.02	41.16	76.80	78.72
Jogging	22.48	95.44	21.50	22.48	21.17	22.48	96.74	21.50	95.11
Liquor	20.85	20.96	22.92	20.10	32.45	40.61	56.17	57.96	69.79

fail. Our local tracking model draws the visible part and keeps the track. The Basketball sequence has many deformations, but we still track accurately in the end.

Out of Plane Rotation: Tracking target rotation is also a big challenge in the field of visual tracking. In Fig. 2(e), the object rotates 1/4 turn. More



Fig. 2. Representative frames from ten sequences. The results obtained by those ten state-of-the-art algorithms and ours are shown in different colors: MIL in pink, VTD in purple, CT in green, DF in black, SCM in gray, CXT in blue, TLD in turquoise, Struck in orange, STC in dark red, ONNDL in cyan, and Ours in red. (Color figure online)

than half of trackers cannot handle with the situation, but our algorithm can implement accurate tracking.

Background Clutter: In the four background clutter sequences (Basketball, David3, Football and Liquor), our tracker performs more stable than other trackers. In the Basketball and Football sequences, there are many players wearing the same clothes. The background near the target has the similar color or texture as the target in the David3 and Liquor sequence. Background clutter can lead to drafting. However, our method achieves better tracking performance.

Both table and figures show that our method achieves favorable performance against other state-of-the-art methods.

4 Conclusion

In this paper, we present a new visual tracking algorithm based on ensemble learning using logistic regression model. The sample is represented by haar-like features. The logistic regression model is adopted to obtain the weights of weak classifiers. The selection of weak classifier and weights of classifiers are implemented simultaneously. The experimental results show the effectiveness of the proposed method.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grant 61571342, 61573267, 61473215; by the National Basic Research Program of China under Grant 2013CB329402.

References

1. Biederman, I., Subramaniam, S., Bar, M., et al.: Subordinate-level object classification reexamined. *Psychol. Res.* **62**(2–3), 131–153 (1999)
2. Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., Belongie, S.: Visual recognition with humans in the loop. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010*. LNCS, vol. 6314, pp. 438–451. Springer, Berlin (2010). doi:[10.1007/978-3-642-15561-1_32](https://doi.org/10.1007/978-3-642-15561-1_32)
3. Hillel, A., Weinshall, D.: Subordinate class recognition using relational object models. In: *NIPS*, pp. 73–80 (2006)
4. Yang, J., Yu, K., Gong, Y., et al.: Linear spatial pyramid matching using sparse coding for image classification. In: *CVPR*, pp. 1794–1801 (2009)
5. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: *CVPR*, pp. 1470–1478 (2003)
6. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. In: *BMVC*, pp. 23.1–23.11 (2009)
7. Yao, B.B., Bradski, G., Li, F.F.: A codebook-free and annotation-free approach for fine-grained image categorization. In: *CVPR*, pp. 3466–3473 (2012)
8. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: *CVPR*, pp. 2169–2178 (2006)
9. Sánchez, J., Perronnin, F., Mensink, T.: Image classification with the Fisher vector: theory and practice. *Int. J. Comput. Vis.* **105**(3), 222–245 (2013)

10. Perronnin, F., Sánchez, J., Mensink, T.: Improving the Fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 143–156. Springer, Berlin (2010). doi:[10.1007/978-3-642-15561-1_11](https://doi.org/10.1007/978-3-642-15561-1_11)
11. Perronnin, F., Dance, C.: Fisher kernels on visual vocabularies for image categorization. In: CVPR, pp. 1–8 (2007)
12. Zhang, J., Marszalek, M., Lazebnik, S., et al.: Local features and kernels for classification of texture and object categories: a comprehensive study. *Int. J. Comput. Vis.* **73**(2), 213–238 (2005)
13. Liu, H., Su, Z.: Template-based multiple codebooks generation for fine-grained shopping classification, retrieval. In: ICDH, pp. 293–298 (2014)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
15. Van de Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1582–1596 (2010)
16. Hiremath, P.S., Pujari, J.: Content based image retrieval using color, texture, shape features. In: ADCOM, pp. 780–784 (2007)
17. Yu, J., Qin, Z., Wan, T., et al.: Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing* **120**, 355–364 (2013)
18. Li, L.J., Su, H., Xing, E., Li, F.F.: Object bank: a high-level image representation for scene classification and semantic feature sparsification. In: NIPS, vol. 26(6), pp. 719–729 (2010)
19. Maji, S., Bourdev, L., Malik, J.: Action recognition from a distributed representation of pose, appearance. In: CVPR, pp. 3177–3184 (2011)
20. Coates, A., Lee, H.: An analysis of single-layer networks in unsupervised feature learning. In: AISTATS, pp. 215–233 (2011)
21. Welinder, P., Branson, S., Mita, T., et al.: Caltech-UCSD birds 200. Technical report, Caltech (2010)
22. Farrell, R., Oza, O., Zhang, N., et al.: Birdlets: subordinate categorization using volumetric primitives and pose-normalized appearance. In: ICCV, pp. 809–818 (2011)
23. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: CVPR, pp. 2169–2178 (2006)
24. Yao, B.B., Khosla, A., Li, F.F.: Combining randomization, discrimination for fine-grained image categorization. In: CVPR, pp. 1577–1584 (2011)