Pattern Recognition 60 (2016) 793–801

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Object detection using boosted local binaries

Haoyu Ren*, Ze-Nian Li

Vision and Media Lab, School of Computing Science, Simon Fraser University, 8888 University Drive Burnaby, BC, Canada V5A1S6

ARTICLE INFO

Article history:
Received 23 October 2014
Received in revised form
4 July 2016
Accepted 4 July 2016
Available online 5 July 2016

Keywords:
Binary descriptor
Boosted Local Binary
Object detection
RealAdaBoost
Structure-aware

ABSTRACT

This paper presents a novel binary descriptor Boosted Local Binary (BLB) for object detection. The proposed descriptor encodes variable local neighbour regions in different scales and locations. Each region pair of the proposed descriptor is selected by the RealAdaBoost algorithm with a penalty term on the structural diversity. As a result, confident features that are good at describing specific characteristics will be chosen. Moreover, the encoding scheme is applied in the gradient domain in addition to the intensity domain, which is complementary to standard binary descriptors. The proposed method was tested using three benchmark object detection datasets, the CalTech pedestrian dataset, the FDDB face dataset, and the PASCAL VOC 2007 dataset. Experimental results demonstrate that the detection accuracy of the proposed BLB clearly outperforms traditional binary descriptors. It also achieves comparable performance with some state-of-the-art algorithms.

1. Introduction

Object detection is one of the most important tasks in computer vision. It is widely used in human–computer interaction, multimedia application, and medical imaging. This task is relatively difficult because there are numerous factors affecting the performance, such as illumination variation, occlusion, as well as background clutters. All these factors will increase the difficulty of classifying the target object with surrounding backgrounds.

To solve this issue, some researchers focus on designing effective descriptors, e.g., Haar [1], Histogram of Oriented Gradient (HOG) [2], covariance matrix [3] or their combinations, e.g., heterogeneous feature [4], HOG-LBP [5], feature fusion [6]. Others utilize more powerful machine learning algorithms, e.g., latent SVM [7], multiple instance learning [8], and Hough forest [9]. Among these algorithms, the cascade boosted structure proposed by Viola and Jones has shown its efficiency and effectiveness on object categories such as faces [1], pedestrians [10] and cars [11]. In most of the cases, the performance of the boosted classifier mainly depends on the features. To address the accuracy and efficiency issues simultaneously, boosting with appropriate features to construct the cascade classifier is the key step.

In consideration of the efficiency, binary descriptors are one of the most commonly used descriptors in object detection. The Local Binary Pattern (LBP) [12] is a local descriptor based on binary coding of adjacent pixel pairs, which is widely used for many pattern recognition tasks. Despite its simplicity, a number of LBP modifications and extensions have been proposed. Some of them work on the post-processing steps [13] or the co-occurrence structure [14] which improve the discriminative ability of binary coding. Others focus on the definition of the location where the gray value measurement is taken [15,16]. Unfortunately, these descriptors have drawbacks when employed to encode general object's appearance. A notable disadvantage is the insufficient discriminative ability. Most of the traditional binary descriptors depend on the intensities of particular locations. It will be easily influenced by illumination, occlusion, and noises. In addition, although the size of some binary descriptors are flexible, the patterns of the local pixels and adjacent rectangles are still fixed. It might not have sufficient ability to describe the objects in some complicate detection tasks, e.g, pedestrians and multi-view cars.

This paper is an extension of our previous work [17] with the following contributions. First, we introduce a Boosted Local Binary (BLB) descriptor, where the variable local region pairs are selected by the RealAdaBoost algorithm considering both the discriminative ability and the feature structural diversity. In addition, we show that using the gradient image and intensity image together for binary coding is more effective than only using the intensity image. As a result, the proposed BLB descriptor is more discriminative and robust compared to commonly used binary descriptors such as Haar, LBP and LAB. We evaluate the performance by employing it on three commonly used datasets, the FDDB face dataset, CalTech Pedestrian dataset and PASCAL VOC 2007 dataset. Experimental results show that BLB has a superior performance in comparison with traditional binary descriptors.

* Corresponding author.
E-mail addresses: hra15@sfu.ca (H. Ren), li@sfu.ca (Z.-N. Li).

http://dx.doi.org/10.1016/j.patcog.2016.07.010
0031-3203/

The rest of the paper is organized as follows: related work will be introduced in Section 2. Section 3 gives the details of the proposed BLB. The structure-aware RealAdaBoost for BLB is described in Section 4. The next section presents the experimental results. Conclusions are given in Section 6.

2. Related work

There have been a wide variety of approaches developed for object detection. Most of them focus on designing more discriminative local descriptors and using appropriate machine learning methods.

There are many local features and descriptors proposed for various detection tasks. Most of them reflect the characteristic of some pre-defined local patterns. The template descriptors which are based on intensity values are widely used for object detection [1,12,16,18–25]. Among these template features, Local Binary Pattern (LBP) is widely used. After the pioneering work [12], Tan and Triggs [18] propose a local ternary pattern (LTP) for robust face recognition. Yan et al. [16] utilize rectangles instead of the single pixel to generate the Local Assembled Binary (LAB) for face detection. Guo et al. [19] propose a weighted LBP method, where the variance that characterizes the local contrast information is used to weight the one dimensional LBP histogram. Vu and Caplier propose oriented edge patterns [20] and patterns of dominant orientations [22] for face recognition. Guo and Zhang [21] propose a completed LBP (CLBP) feature to incorporate the sign and magnitude information into the final descriptor. Ahonen et al. [24] propose the LBP Fourier histogram (LBPHF) to achieve rotation invariance. Zhao et al. [25] propose an extension of LBPHF, by combining the sign and magnitude information. Although these LBP mutations improve the accuracy considerably compare to traditional LBP, they do not have good generalization power due to the artificially designed local patterns. For some general object detection tasks such as PASCAL VOC challenge, such that the object appearance varies a lot with complex background, these features will not work well.

Besides the binary descriptors, more complicate features and descriptors have been utilized, such as SIFT [26], HOG [2], and covariance matrix [3]. Lowe [26] first proposes the scale invariant feature transform (SIFT). Several mutations are designed in these years for object detection and recognition [27–30]. Dalal and Triggs [2] propose the basic form of the HOG descriptor with 2×2 cells. Multi-size version are developed in [10,31,32], and further extended to pyramid structure [33–35]. Tuzel et al. [3] utilize the covariance matrix projected on Riemann manifolds for detection. A heterogeneous version based on covariance matrix is further proposed in [36]. Sometimes these features are combined with each other to increase the discriminative power. For instance, Levi and Silberstein [37] utilize an accelerated version of the feature synthesis method applied on multiple object parts respectively. Bar-Hillel et al. [38] design an iterative process including feature generation and pruning using multiple operators for part localization. Chen et al. [39] propose Multi-Order Contextual co-Occurrence (MOCO), to implicitly model the high level context using solely detection responses from the object detection based on the combination of HOG and LBP. Using complicate features clearly improves the accuracy compared to binary descriptors, but the efficiency is reduced at a large scale. Most of these features could not satisfy the requirement of real-time object detection systems. As a result, the development of effective binary descriptors is necessary.

Boosting framework is widely used in training the cascade classifier for fast object detection. Zhu et al. [10] apply linear SVM with HOG descriptor as the weak classifier to build a cascade detector. This procedure is revisited through properly designing the feature pooling, feature selection, preprocessing, and training methods using a single rigid component [40]. Wu and Nevatia [41] propose the cluster boosted tree method, in which the sample space is divided by unsupervised clustering based on discriminative image features selected by boosting algorithm. Tu [42] develops the probabilistic boosting-tree, where each node combines a number of weak classifiers (evidence, knowledge) into a strong classifier (a conditional posterior probability).

3. Boosted binary patterns

3.1. Traditional binary descriptors

The traditional LBP is developed for texture classification and the success is due to its robustness under illumination variations, computational simplicity and discriminative power on specific patterns. Fig. 1 represents an example of the traditional LBP, which is a binary coding of the intensity contrast of the center pixel and 8 neighbouring pixels. If the intensity of neighbouring pixels are higher than the center one, the corresponding bits will be assigned 1, otherwise it will be assigned 0. Given a center pixel, the LBP feature response is defined by

$$LBP_{d,r} = \sum_{i=1}^{d} \text{sign}(I_i - I_{center}) \times 2^{i-1}, \tag{1}$$

where d is the number of neighbouring pixels, r is the distance between the neighbouring pixels and the center pixel, I is the intensity, and

$$\text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

Different from LBP which reflects the intensity pattern of pixel pairs, LAB [16] utilizes rectangles instead, as shown in Fig. 2. LAB combines 8 locally adjacent 2-rectangle binary Haar features with the same size. These Haar features share a common center rectangle. LAB's encoding scheme is similar to LBP: if the intensity sum of the adjacent rectangle is higher than the center one, the corresponding bit will be assigned 1; otherwise it will be assigned 0. Given a center rectangle C_0, the LAB feature response is

$$LAB = \sum_{i=1}^{8} \text{sign}\left(I_{C_i} - I_{C_0}\right) \times 2^{i-1}, \tag{2}$$

where C_i is the adjacent rectangle, I_{C_i} is the intensity sum of all the pixels in C_i.

The calculation of LBP is efficient because the feature response in Eq. (1) is based on binary comparisons and bit-shift operations. Although LAB uses regions instead of pixels, the computation cost will not increase because the intensity sum of any rectangles could be efficiently calculated using the integral image [1].

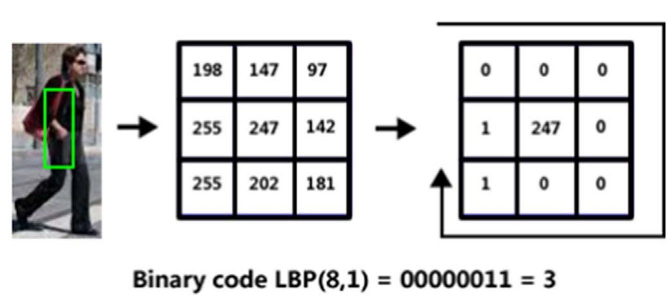

Fig. 1. Traditional LBP descriptor.