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 FDB 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 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 complicated 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 FDB 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.
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), by combining the sign and magnitude information. Although these LBP mutations improve the accuracy considerably compared to the 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 complicated 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 Silberman [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

\[ \text{LBP}_{r,x} = \sum_{i=1}^{d} \text{sign}(k - l_{x}) \times 2^{i-1}, \]

where \( d \) is the number of neighbouring pixels, \( r \) is the distance between the neighbouring pixels and the center pixel, \( l \) 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

\[ \text{LAB} = \sum_{i=1}^{8} \text{sign}(l_{i} - l_{0}) \times 2^{i-1}, \]

where \( l_{i} \) is the adjacent rectangle, \( l_{0} \) is the intensity sum of all the pixels in \( C_0 \).

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].
3.2. Boosted local binaries

Instead of using the artificially designed binary patterns, we introduce the proposed Boosted Local Binaries (BLB) descriptor. A BLB-n descriptor consists of \( n \) non-overlapped surrounding rectangles \( C_0, C_1, ..., C_n \) and a center rectangle \( C_0 \) in the same size. As in Eq. (3), the \((x_0, y_0)\) is the left-top corner of the rectangle \( C_i \), and \( (w_i, h_i) \) is the width and height

\[
BLB_n = \{ C_0, C_1, ..., C_n \}
\]

\[
C_i = (x_i, y_i, w_i, h_i), \quad w_i = w_j, \quad h_i = h_j, \quad i, j \in \{0, 1, ..., n\}, \quad i \neq j. \tag{3}
\]

These rectangles compose \( n \) pairs \( \{C_0, C_i\}, i = 1, ..., n \). In our case, the \( n \) could be 4, 6, or 8. Fig. 3 illustrates the examples of BLB-4, BLB-6, BLB-8 descriptors.

In BLB, the surrounding rectangles \( C_0, C_1, ..., C_n \) are numbered clockwise starting from the one at the top of the center rectangle \( C_0 \). None of these surrounding rectangles are overlapped with each other. Fig. 4(a) shows a BLB-8 descriptor, where the center rectangle \( C_0 \), top rectangle \( C_1 \), right-top rectangle \( C_2 \) and left-top rectangle \( C_8 \) are highlighted. In general, the surrounding rectangles \( C_0, C_1, C_8 \) could spread anywhere. To describe meaningful patterns and reduce the size of the feature pool, we define the neighbouring pairs illustrated as the rectangles connected by green lines in Fig. 5. There are 4 neighbouring pairs in BLB-4, 12 in BLB-6, and 12 in BLB-8. We add a constraint that the two rectangles in each neighbouring pair should overlap either on their width or on their height. The overlap ratio of BLB-4, BLB-6, and BLB-8 is set to 0.5, 0.25, and 0.5 respectively. Based on this rule, the \( C_i \) in Fig. 4(a) should overlap at least 50% with \( C_0, C_2, C_8 \) either on their width or on their height. This is illustrated by the dot-dashed lines around \( C_0, C_2, C_8 \) which reflect the possible region of their neighbours. As a result, the possible region for \( C_1 \) is the cyan region which is intersected by the corresponding dot-dashed lines of \( C_0, C_2 \) and \( C_8 \). Similarly, Fig. 4(b) illustrates a BLB-6 descriptor. \( C_1 \) is included in three neighbouring pairs \( \{C_0, C_1\}, \{C_0, C_2\}, \{C_2, C_1\} \). Then the possible region for \( C_1 \) is the green region intersected by the dot-dashed lines.

If the surrounding rectangles lay far away from the center one, the feature response will be easily influenced by noises. As a result, the classification confidence of this descriptor is lower. We define the structural diversity as the average ratio of the distance between each surrounding rectangle and the center rectangle to the rectangle size as

\[
D = \frac{1}{n} \sum_{i=1}^{n} \frac{2 \text{dist}(C_i, C_0)}{w_i + h_i}. \tag{4}
\]

where the \( \text{dist}(x, y) \) is the Euclidean distance between \( x \) and \( y \). If the surrounding rectangles are far away from the center one, \( D \) will be larger. As shown in Fig. 6, the surrounding regions of the left BLB-6 descriptor are closer to the center one, so its diversity is smaller. In contrast, the surrounding rectangles of the right BLB-8 descriptor spread farther, so that the diversity is relatively larger.

![Fig. 2. LAB descriptor.](image1)

![Fig. 3. BLB descriptors. The black rectangle is the center rectangle, and the white ones are the surrounding rectangles.](image2)

![Fig. 4. Constraint of the neighbouring pairs. (a) For BLB-8, the overlap ratio is 50%. (b) For BLB-6, the overlap ratio is 25%. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)](image3)

![Fig. 5. Neighbouring pairs in BLB-4, BLB-6, and BLB-8. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)](image4)

![Fig. 6. Structural diversity of BLB descriptor. The left BLB-6 has smaller diversity, while the right BLB-8 has a larger one.](image5)
This diversity is designed to reflect the BLB’s structural confidence. It will be considered as a penalty term in the RealAdaBoost procedure.

Besides applying the above patterns on the intensity image, we also utilize them on gradient images. In consideration of the efficiency, the x-direction gradient image and y-direction gradient image are generated respectively. Given a BLB-n descriptor, the final feature response is

$$f(\text{BLB}_n, \text{id}x) = \sum_{i=1}^{n} \text{sign}(g(C_i, \text{id}x) - g(C_{\text{ref}}, \text{id}x)) \times 2^{i-1},$$

where

$$g(C_i, \text{id}x) = \begin{cases} \text{IntensitySum}(C_i) & \text{id}x = 0 \\ \text{GradientxSum}(C_i) & \text{id}x = 1 \\ \text{GradientySum}(C_i) & \text{id}x = 2 \end{cases}$$

Compared to conventional LBP or LAB, the proposed BLB has clear advantages. We know that if we average several pedestrians in the intensity domain, the result will be meaningless. On the other hand, if we average them in the gradient domain, we may easily find some common characteristic. This indicates the gradient information is quite helpful in detecting some object categories. In addition, the variable binary patterns are effective in capturing some specific object structure which could not be described, such as the contrast of window, body, and wheel in a car. Moreover, the computation of BLB is as efficient as LBP and LAB. The sum of a region in both intensity image and gradient image could be easily extracted by the integral images [1]. So the efficiency is similar to traditional binary descriptors.

### 4. RealAdaBoost with BLB

We use RealAdaBoost to select the meaningful BLB descriptors to describe the target object. In RealAdaBoost, the BLB descriptor can be seen as a function from the image space to a real valued range $f: \mathbf{x} \rightarrow [f_{\text{min}}, f_{\text{max}}]$. For the binary object/background classification problem, suppose the input data as $(\mathbf{x}_i, y_i), \ldots, (\mathbf{x}_N, y_N)$ where $\mathbf{x}_i$ is the training sample and $y_i \in \{-1, 1\}$ is the class label, we first divide the sample space into $N_b$ equal sized sub-ranges $B_j$ and the weak classifier is defined as a piecewise function:

$$h(\mathbf{x}) = \frac{1}{2} \ln \left( \frac{W_+ + \epsilon}{W_- + \epsilon} \right),$$

where $\epsilon$ is the smoothing factor, $W_+$ are the probability distributions of the feature response for positive/negative samples, implemented as a histogram

$$W_j = P(\mathbf{x} \in X_j, y \in \{-1, 1\}) \quad j = 1, \ldots, N_b.$$

The best descriptor is selected according to the classification error $Z$ of the piecewise function

$$Z = 2 \sum_j \sqrt{W_j W_j} + a \cdot \text{fp} \cdot D.$$  

Eq. (10) could be explained as follows: in the beginning stages of RealAdaBoost, because the false positive rate is larger, and the target object is still easy to be classified with the background, RealAdaBoost will refer to the smaller-diversity descriptors with more confident patterns. In the following stages when the false positive rate is smaller, the classification problem becomes more difficult, so descriptors with diverse patterns might be utilized. This strategy makes sense, because the overall performance and robustness of a cascade boosted classifier is strongly influenced by the beginning stages which filter most of the candidate windows. Using confident features in the beginning stage will contribute to the generalization power of the resulting classifier, which is important in real object detection system.

The detail algorithm is illustrated in Fig. 7. To learn the best feature, the most intuitive way is to look through the whole feature pool, which is rather time consuming. So we resort to a sampling method to speed up the feature selection process. Both of the sampling number $M$ and bin number $N_b$ will influence the total performance. Further experiments on these parameters are given in Section 5.1.
there is a signiﬁcant intensity information for BLB-8 and BLB-All. It can be seen that of resulting detector. We also train two classiﬁers to further improved around 2%. This shows that the variable BLB descriptors are integrated together, the accuracy could be proved along with the increase of BLB surrounding rectangles. If all FPPI of these classiﬁers are able to re-extract the above descriptors in both the gradient domain (gra) and the intensity domain (int). Fig. 8(a) shows the miss rate versus FPPI (False Positive Per Image) curves are utilized to evaluate the performance of our algorithms. 64 × 128 samples are used in all the experiments on this dataset. BLB descriptors with variable block size from 4 × 4 to 20 × 40 are evaluated each iteration.

The next experiment is about the sampled feature number in order to compare our method with traditional binary descriptors and other state-of-the-art algorithms, we use the optimal parameters (BLB-All, 60 features per iteration, structure-aware process). From Fig. 9(b) we could ﬁnd that using 64 bins achieves similar accuracy compared to using 32 bins. More bins such as 128 does not increase the accuracy much. This is another trade-off problem. More bins means trying to achieve more accuracy, and less robustness. As a result, it leads to higher risk to generate more false positives, especially in the large database. Using the structure-aware process to balance them could improve the overall performance. The best one, C=0.2, contributes to almost 5% accuracy improvement compared to the classiﬁer without structure-aware process. We know that in image-based object detection, the total accuracy is decided not only by the discriminative ability of the detector, but also by the generalization power. Although the BLB descriptors with larger structural diversity might capture some complicate characteristics, they are also easily inﬂuenced by noises. As a result, it leads to higher risk to generate more false positives.

In order to evaluate the difference between the pedestrian and background. The advantage of using gradient information is clear.

The next experiment is about the sampled feature number in each iteration. 4 classiﬁers are trained by setting the sampled feature number to 30, 60, 200, and 500. All BLB descriptors are utilized, and the RealAdaBoost bin number is set to 32. From Fig. 8(b) we can ﬁnd that there is a clear blank between the 30 and 60 curve. In addition, evaluating 200 or 500 features per iteration only leads to slightly improvement. Compared to the increased training cost, this improvement is not valuable. According to [10], a random sub-sample of size \( N_f = \log 0.05/\log 0.95 = 59 \) will guarantee that we can ﬁnd the best 5% features with a probability of 95%. So using 60 features per iteration ensures us get a detector with the top-level features.

Then the inﬂuence of the structure-aware process is evaluated. Based on previous experiments, 60 BLB descriptors are sampled per iteration. The RealAdaBoost bin number is ﬁxed at 32. Fig. 9(a) gives several curves based on different structure-aware factors C. It can be seen that the structure-aware process contributes to at least 2% improvement on the detection accuracy. The best one, C=0.2, contributes to almost 5% accuracy improvement compared to the classiﬁer without structure-aware process. We know that in image-based object detection, the total accuracy is decided not only by the discriminative ability of the detector, but also by the generalization power. Although the BLB descriptors with larger structural diversity might capture some complicate characteristics, they are also easily inﬂuenced by noises. As a result, it leads to higher risk to generate more false positives, especially in the large database. Using the structure-aware process to balance them could improve the overall performance of the resulting classiﬁer. This is similar to the efﬁciency-accuracy trade-off in some sense.

The last parameter is the RealAdaBoost bin number \( N_b \). From Fig. 9(b) we could ﬁnd that using 64 bins achieves similar accuracy compared to using 32 bins. More bins such as 128 does not increase the accuracy much. This is another trade-off problem. More bins means ﬁner division of the feature space, which leads to better discriminative power but less robustness. On the other hand, fewer bins will do harm to the overall accuracy, as the 2.4% gap between 16 bins curve and 32 bins curve. Based on our experiments, 32 bins seems to be a good choice for BLB.

In order to compare our method with traditional binary descriptors and other state-of-the-art algorithms, we use the optimal parameters (BLB-All, 60 features per iteration, structure-aware factor 0.2, 32 RealAdaBoost bins) to train a ﬁnal classiﬁer. When the false positive rate decreases to \( 10^{-6} \), 6671 BLB descriptors are utilized. 4117 of them are BLB-8, 1701 are BLB-6, while the remaining parts are BLB-4. 2739 of these 6671 descriptors are extracted on the intensity domain, 1871 are in x-gradient domain,

**Fig. 8.** Experiments on different BLB descriptors. (a) Different number of surrounding rectangles and feature extraction methods. (b) Different sampled feature number.
2061 are in y-gradient domain. The LBP and LAB descriptors are also implemented. For LBP, 3 kinds of LBP descriptors, LBP(8,1), LBP(8,2), and LBP(16,2) are integrated together to train a cascade detector. For LAB descriptor, variable block size from 4 to 20 are used. We also evaluate 60 descriptors per iteration and set the bin number to 32 for these two descriptors. In Fig. 10, we illustrated the results of traditional Haar feature, LBP(8,1) feature, the combination of all LBP features, and LAB feature. The “int” curves show that this feature is only extracted on the intensity image, while the “int+gra” curves indicate that the features are extracted from both the intensity and gradient domain. It could be seen that the best group among these traditional features, LBP(int+gra) achieves 53% miss rate, which is still 6% lower compared to the proposed BLB (47%). This result shows the effectiveness of our variable patterns compared to fixed patterns. In addition, using more patterns for LBP (All LBP) achieves better accuracy compared to single pattern (LBP(8,1)). This indicates the advantage of variable binary patterns. We also notice that using both the intensity and gradient information is more effective compared to only using the intensity information. This indicates that the gradient characteristic is important for the pedestrian detection.

Moreover, we compare the convergence speed of the training process of different descriptors. Fig. 12 plots the curves of false positive rate against the number of weak classifiers. The RealAdaBoost without structure-aware process is utilized in this experiment. It shows that no matter the gradient information is use or not, the BLB always converges faster, at the rate of approximately two times compared to LAB and LBP. We also notice that using gradient information contributes to a significantly faster convergence speed on both LBP, LAB, and BLB. So both the variable feature patterns and the gradient information improve the discriminative ability. In addition, the performance of the boosted classifier is shown to be positively proportional to the convergence speed in the training process. This signifies that the proposed BLB performs better on the training accuracy and the training speed of boosted classifiers. We test the resulting classifier on a desktop PC with 2.8 GHz dual core CPU and 8 GB memory. It only takes 17 ms to detect all pedestrians in a 640×480 input image while the minimum pedestrian size is 30×60. Fig. 13(a) shows the first
selected two descriptors of the pedestrians. It clearly captures the body structures of the pedestrian.

### 5.2. Experiments on FDDB dataset

Next, we tested the detectors on the publicly available FDDB data set [45]. This data set contains 5171 faces in 2845 images. Most images include face in various pose, variable size and complicated background. We collect 2400 faces from the website and resize them to $24 \times 24$ as training images. Variable size BLB descriptors from $2 \times 2$ to $6 \times 6$ block size are generated and further selected using the algorithm in Fig. 7. Fig. 14 plots the curves of our method as well as popular binary descriptors, including Haar, LBP, LAB, and the state-of-the-art algorithms including [46–53], in terms of the number of false positives with respect to the detection rate. As shown in Fig. 14, our detector achieves 85.4% detection rate at 200 false positive. To our knowledge, it is the best result for 200 false positive on this dataset. The performance of our method is also similar to the state-of-the-art algorithms with more complicated classification algorithms such as joint cascade [52] or deformable part model [51]. Different from general object detection, there are more binary characteristics in human face, such as the eyebrow to the forehead, or the pupil to the white of eye. These binary relationships also exist even in multi-view faces. Unfortunately, it is difficult for traditional binary descriptors such as Haar or LBP to capture these relationships due to their limited feature patterns. In contrast, BLB successfully captures these meaningful patterns, as illustrated in Fig. 13(b), the contrast of the eyes to the surrounding parts are well reflected in the selected two BLBs.

5.3. Experiments on PASCAL VOC 2007 dataset

Finally, we employed the standard benchmark object detection dataset, PASCAL VOC 2007, to test our detector. The PASCAL dataset contains images from 20 different categories with around 5000 images for training and validation and a test set of size around 5000. We measure the object detection performance on the test images using the standard protocol: average precision (AP) per class, as well as the mean AP (mAP) across all classes. For both measures, we consider that a window is correct if it has an intersection-over-union ratio of at least 50% with a ground-truth object instance.

The sample size is different for these object categories in our training process. For the aeroplane, bird, bottle, chair, diningtable, person, pottedplant, sofa, and TV monitor categories, all the samples are placed together to train a single detector. For all other categories, the training samples are divided into the front/rear view samples and side-view samples, and then trained into two detectors respectively. The final detection result is based on the voting of these two detectors. The sample sizes $(w, h)$ used to train these classifiers are listed in the second column of Table 1. The label ‘f’ means the sample size for front/rear images. Only one detector for this category if there is no ‘f’ label. The size of the BLB rectangles used in these categories ranges from $4 \times 4$ to $w \times h$, where $w$, $h$ are the maximum multiple of 4 smaller than $w/3$ and $h/3$ respectively.

In Table 1, we compare our method with the state-of-the-art algorithms [39,54–56] in terms of detection AP on the test set. It can be seen that the BLB still achieves 5% better mAP compared to LBP and LAB. Firstly, we notice that BLB performs poor on some categories with complicated appearance. For instance, in the bus detection, because the bus bodies are always painted by various advertisements with various colors, there are few meaningful binary patterns to be captured in both the intensity domain and the gradient domain. Although the detector does select a few BLB descriptors which captures the contrast of the window to the bus body, but the classification margins are relatively small, which shows that the detector does not have sufficient confidence of this decision. Similar circumstance happens in motorbike detection and horse detection, where a large number of pictures include the drivers and riders wearing various clothes. For some object categories with simple appearance such as car, BLB works well as complex features such as SIFT fisher vectors [54], pyramid HOG [56], covariance matrix [55] and co-occurrence features [39].

![Fig. 12. Convergence speed of different descriptors.](image1.png)

![Fig. 13. The first selected two descriptors of pedestrians and faces.](image2.png)

![Fig. 14. Experiments on FDDB face database.](image3.png)
addition, if the target has stable structure, such as the potted plant which is always with a base, or the chair which consists of several rigid parts, such that the major characteristics of these objects are binary relationship or contours, BLB is always able to capture these information, which might be ignored by other high-level gradient features. Fortunately, this case does not happen on the aeroplane and sofa. The reason is that in the aeroplane detection, many pictures are truncated, which include only a small part of the aeroplane. In the sofa detection, at least half of the sofa images are occluded by person or other backgrounds. As a result, part-based model [55,56] has advantages dealing with such cases rather than using single boosted classifier.

### 6. Conclusions

In this paper, we propose an object detector that achieves both high accuracy and fast speed. A large binary feature pool with variable-location and variable-size blocks on both the intensity domain and the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different binary relationship or contours, BLB is always able to capture these information, which might be ignored by other high-level gradient features. As a result, the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different and the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different and the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different and the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different and the gradient domain is built. The RealAdaBoost algorithm is able to select informative features by evaluating different

### Acknowledgment

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the Grant RGP36726.

### References


Ze-Nian Li is a Professor in the School of Computing Science at Simon Fraser University, British Columbia, Canada. Dr. Li received his undergraduate education in Electrical Engineering from the University of Science and Technology of China, and M.Sc. and Ph.D. degrees in Computer Sciences from the University of Wisconsin-Madison under the supervision of the late Professor Leonard Uhr. His current research interests include computer vision, multimedia, pattern recognition, image processing, and artificial intelligence. Dr. Li has published over 150 refereed papers in journals and conference proceedings. He is the co-Director of the Vision and Media Lab. He is also the co-author of the book “Fundamentals of Multimedia”, 2nd ed. published by Springer.

Haoyu Ren received his B.Sc. degree from Tsinghua University in 2007, and M.Sc. degree from the Institute of Computing Technology, Chinese Academy of Science in 2010. He worked as the Staff Researcher in the multimedia group of Beijing Samsung Telecom R&D from 2011 to 2013. He is currently a Ph.D. candidate at the School of Computing Science, Simon Fraser University. His research interests include object detection, tracking and recognition based on machine learning methods.