Haar Wavelets and Edge Orientation Histograms for On–Board Pedestrian Detection

David Gerónimo, Antonio López, Daniel Ponsa, and Angel D. Sappa

Computer Vision Center, Universitat Autònoma de Barcelona
Edifici O, 08193 Bellaterra, Barcelona, Spain
{dgeronimo,antonio,daniel,asappa}@cvc.uab.es
www.cvc.uab.es/adas

Abstract. On–board pedestrian detection is a key task in advanced driver assistance systems. It involves dealing with aspect–changing objects in cluttered environments, and working in a wide range of distances, and often relies on a classification step that labels image regions of interest as pedestrians or non–pedestrians. The performance of this classifier is a crucial issue since it represents the most important part of the detection system, thus building a good classifier in terms of false alarms, missdetection rate and processing time is decisive. In this paper, a pedestrian classifier based on Haar wavelets and edge orientation histograms (HW+EOH) with AdaBoost is compared with the current state–of–the–art best human–based classifier: support vector machines using histograms of oriented gradients (HOG). The results show that HW+EOH classifier achieves comparable false alarms/missdetections tradeoffs but at much lower processing time than HOG.

1 Introduction

On–board pedestrian detection in the context of advanced driver assistance systems (ADAS) has become an active research field aimed at reducing the number of traffic accidents. The objective is to provide information to the driver and to perform evasive or braking actions on the host vehicle by detecting people in a given range of distances. The most relevant works in the literature [1,2,3] base detection on a classification step that labels regions of interest in the input image as pedestrians or non–pedestrians. The main difficulty of the classification stage comes from dealing with aspect–changing targets like pedestrians, which are subject to a high intra–class variability. In Fig. 1 some pedestrian samples illustrate the variability of this object class for different distances, backgrounds, illuminations, poses or clothes.

In this paper we compare two relevant pedestrian classifiers in the context of ADAS. The first one uses Haar wavelets and edge orientation histograms (HW+EOH) as features and Real AdaBoost as learning machine. This classifier is originally proposed by *Levi and Weiss* to perform face detection in [4]. In this paper we add some slight modifications and use it to classify pedestrian samples. In order to evaluate the performance of the mentioned classifier,

J. Martí et al. (Eds.): IbPRIA 2007, Part I, LNCS 4477, pp. 418–425, 2007.

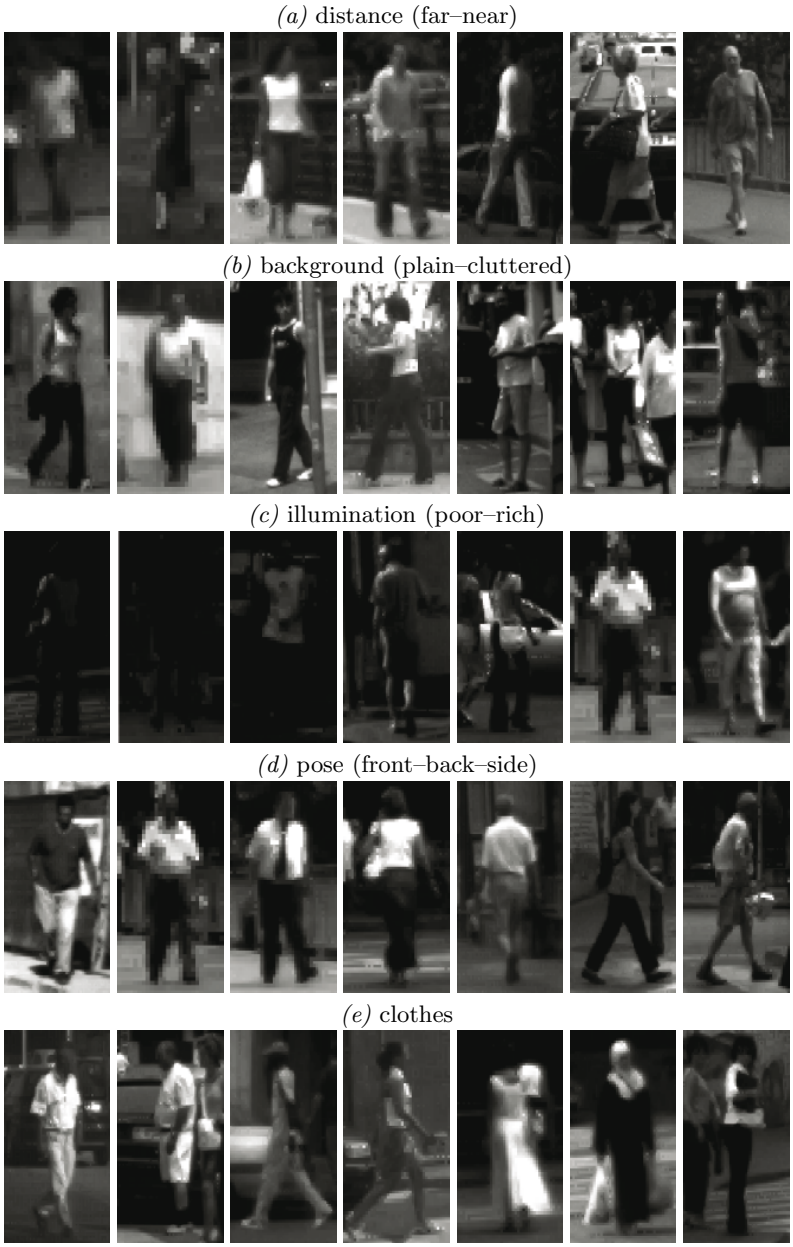

Fig. 1. Positive samples of the database illustrating the high variability in terms of distance, background, illumination, pose and clothes (contrast enhaced for better visualization). Distance variations are specially relevant in ADAS databases. For instance, in this case the sample size can range from 12×24 to 120×240 pixels.

we make a comparison against the best human–detector in the state–of–the–art literature: support vector machines using histograms of oriented gradients (HOG) by *Dalal and Triggs* [5]. To make a relevant comparison, we first tune the feature parameters for an ADAS pedestrian database, selecting the ones that achieve the best performance, and then we analyse the classifiers in terms of false alarms/missdetection rates and processing time. Experiments show that the HW+EOH based classifier achieves the same detection rates than the HOG based one, but being ten times faster. Moreover, HOG rates are outperformed by increasing the complexity of the HW+EOH classifier, but still requiring half the processing time.

The reminder of the paper is organized as follows. Sect. 2 describes the HW+EOH based classifier and Sect. 3 the HOG based one. Sect. 4 presents the database used, and then evaluates the two classifiers in terms of detection rates and processing time. Finally, Sect. 5 exposes the main conclusions.

2 HW+EOH Based Classifier

Levi and Weiss [4] propose a combination of two sets of features, Haar wavelets (HW) and edge orientation histograms (EOH), to detect faces, outperforming the detection rates of the single sets alone. In this paper, we add some modifications in order to improve the detection results: we use Real AdaBoost [6] instead of the original AdaBoost version, and make some slight modifications when computing the features, as described in this section.

Haar wavelets represent a fast and simple way to calculate region derivatives at different scales by means of computing the average intensities of concrete subregions (defined by a set of filters). They are proposed by *Papageorgiou et al.* [3] for object recognition. A feature of this set is defined as the difference of intensity between two defined areas (white and black) in a given position inside a region R:

$$\text{Feature}_{Haar}(x, y, w, h, type, R) = E_{white}(R) - E_{black}(R) \ ,$$

where (x, y) is the bottom–left corner of the filter; w, h are the filter's width and height; and *type* corresponds to the filter's configuration. $E_{white}(R)$ and $E_{black}(R)$ represent the sum of intensities of white and black areas of the template respectively. In order to compute E, the *integral image* (*ii*) representation [7] has been used, where the summed values of a certain region can be efficiently computed by four *ii* accesses.

The original set of filters contains three basic configurations [3] (Fig. 2 (*middle*) (a–c)), that capture changes in intensity along the horizontal, vertical directions and the diagonals. In our case, we use the set proposed by *Viola and Jones* [7], which contains two additional filters (Fig. 2(*middle*)(a–e)). In addition, in this work we have also followed the latter approach [7], where filters are not constrained to a fixed size, as proposed in [3], but can vary in size and aspect ratio.

Due to perspective, different windows framing a pedestrians can have different sizes, so spatial normalization is required to establish an equivalence between the features computed in each window. To achieve that, it is not necessary to

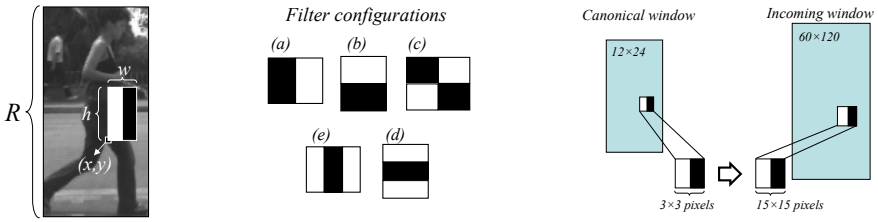

Fig. 2. Computation of Haar wavelet features: ($left$) Haar feature placed in a sample image; ($middle$) some filter configurations; ($right$) filter normalization according to the incoming window size

explicitly resize the windows, but features can be computed in a way that it is equivalent to resizing but more efficient [7]. Our canonical window is 12×24 pixels (Fig. 2($right$)), which in our acquisition system corresponds to a *standard* pedestrian at about 50m. In addition, we modify the filters to be illumination invariant, in order to obtain responses identical to the ones obtained by previously normalizing the contrast of the processed image region.

Edge orientation histograms[1] are also interesting for our work, since pedestrians often present strong edges in the legs or trunk areas. They rely on the richness of edge information, so they differ from the intensity area differences of HW but maintain invariance properties to global illumination changes.

First, the gradient image is computed by a Sobel mask convolution (contrary to the original paper, no edge–thresholding is applied in our case). Then, gradient pixels are classified into β images corresponding to β orientation ranges (also referred as *bins*, in our case we have tested $\beta = \{4, 6, 9\}$). Therefore, a pixel in bin $k_n \in \beta$ contains its gradient magnitude if its orientation is inside β_n's range, otherwise is null. Integral images are now used to store the accumulation image of each of the edge bins.

At this stage a bin interpolation step has been included in order to distribute the gradient value into adjacent bins. This step is used in SIFT [8] and HOG [5] features, and in our experiments the improvement achieved (using EOH features alone) is 1% Detection Rate (DR) at 0.01 False Positive Rate (FPR).

Finally, the feature value is defined as the relation between two orientations, k_1 and k_2, of region R as:

$$\text{Feature}_{EOH}(x, y, w, h, k_1, k_2, R) = \frac{E_{k_1}(R)+\epsilon}{E_{k_2}(R)+\epsilon} \quad .$$

If this value is above a threshold of 1, it can be said that orientation k_1 is dominant to orientation k_2 for R. If the value is lower than 1 it can be said than k_2 is dominant to k_1. The small value ϵ is added for smoothing purposes.

[1] In order to respect the author's work, in this paper we maintain the original name. However, since this can lead to confusion with other similar feature names like the *histograms of oriented gradients* (HOG) in [5] (Sect. 3), we think that a more convenient name would be *ratios of gradient orientations*.

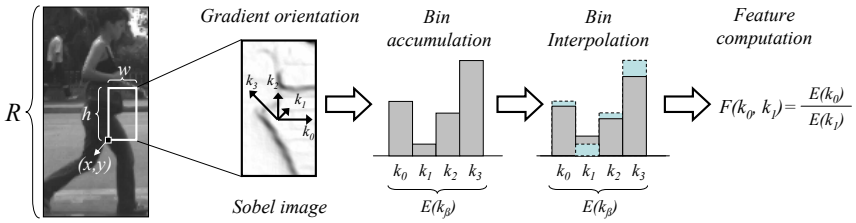

Fig. 3. Computation of edge orientation histograms

In our implementation, we make use of Real AdaBoost [6] as learning machine, rather than the original AdaBoost version used in [4]. The idea is to build a *strong* classifier by combining the response of a set of *weak* classifiers, improving the performance that a complex classifier would have alone. In our case, since both HW and EOH features are represented by a real value, each weak classifier corresponds to a threshold–like rule on each feature value.

In the literature, AdaBoost classifiers are often implemented in a cascade so the number of false positives decreases and hence the overall performance is increased. In this comparison, since we are more interested in the features than in the learning machine, we make use of just one cascade level. However, the multiple–cascade strategy is planned to be implemented in future works.

3 HOG Based Classifier

The current state–of–the–art best classifier is based on histograms of oriented gradients (HOG) as features and support vector machine (SVM) as learning algorithm. It is proposed by *Dalal and Triggs* [5] to perform human detection. HOG are SIFT–inspired features [8] that rely on gradient orientation information. The idea is to divide the image into small regions, named *cells*, that are represented by a 1D histogram of the gradient orientation. Cells are grouped in larger spatial regions called *blocks* so histograms contained in a block are attached and normalized (Fig. 4).

When computing the features, we follow the indications of the authors as strictly as possible. As the authors suggest, no smoothing is aplied to the incoming image, and a simple 1D $[-1, 0, 1]$ mask is used to extract the gradient information. Next, we have tested the best parameters for our database: number of bins ($\beta = \{4, 6, 9\}$ in $0 - 180°$), cell sizes ($\eta = \{1 \times 1, 2 \times 2, 3 \times 3\}$ pixels) and block sizes ($\varsigma = \{1 \times 1, 2 \times 2, 3 \times 3\}$ cells), for our 24×12 canonical windows. Block overlapping is set to the maximum possible, i.e., ς–fold coverage for each cell. Bin interpolation is also used here. As last step, the block histogram is normalized using *L2-Hys*, the best method in the original paper, i.e., L2-normalizing, clipping values above 0.2, and then renormalizing. Finally, the features are fed to a linear SVM (following the authors' indications, SVMLight `<http://svmlight.joachims.org>` with C=0.01 has been used)[2].

[2] Real AdaBoost has also been tested using gradient orientations as weak rules, which results in similar performance rates. Thus, we keep the original formulation.

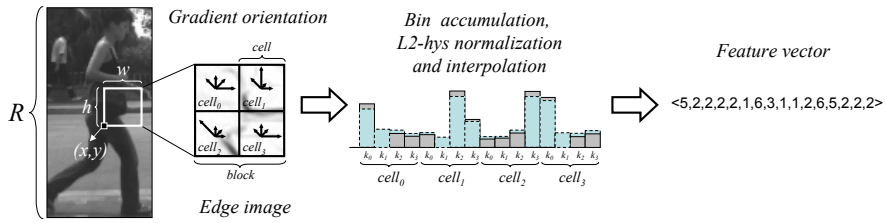

Fig. 4. Computation of histograms of oriented gradients

Although not done in [5], we make use of the integral image representation to store the bins histograms corresponding to each orientation, which dramatically speeds up the features computation. This approach, as previously reported by [9], is incompatible with the Gaussian spatial window applied to the block before constructing the histogram. However, since [9] achieves the same results than [5] without applying the Gaussian spatial window step, we ommit it without danger of significantly decreasing the performance.

4 Experimental Results

In order to illustrate the performance of the classifiers under real driving environments we use our ADAS pedestrian database. Differently to other non ADAS–oriented databases [5], it contains images at different scales from urban scenarios. In our case, color information is discarded as an useful cue, so samples are transformed to grayscale. The complete database consists of $1,000$ positive samples (i.e., pedestrians; Fig. 1) and $5,000$ negative ones (i.e., human–sized windows in regions likely to contain pedestrians). Each experiment randomly selects 700 positive and $4,000$ negative samples (training set) to learn a model, and use the remaining (testing set) to measure the classifier performance. All performance rates and plots are the result of averaging 4 independent experiments.

In order to be rigurous and provide a fair comparison, we have tunned the feature parameters to select the best ones for the database. We have tested $\beta = \{4, 6, 9\}$ for HW+EOH, achieving similar results (Fig. 5($left$)). Hence, we have selected the $\beta = 4$ bins version since it requires less processing time. Regarding to HOG features, the optimum parameters are $\beta = 9$, $\eta = 2 \times 2$ and $\varsigma = 2 \times 2$, which provide a detection rate (DR) of 92.5% at $FPR = 1\%$ (Fig. 5 ($right$)).

Figure 6($right$) presents the comparison between the HW+EOH based classifier and the HOG based one. As can be seen, with 100 features (i.e., Real AdaBoost weak rules) HW+EOH reaches the same performance as HOG. However, the HW+EOH features are at least ten times faster to compute (each window is classified in 0.015 ms). With 500 features the DR improves 4% (at $FPR = 0.01$), and it is computed about two times faster than HOG.

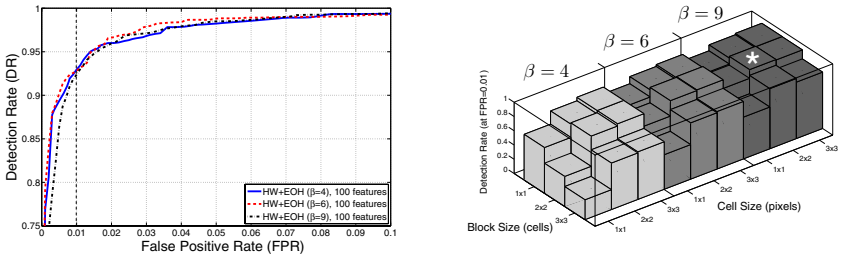

Fig. 5. (*left*) Performance of the proposed classifier using different β for the EOH features. (*right*) Detection rate at FPR=0.01 for all possible configurations of β, η and ς of HOG features (the best one is marked with a star).

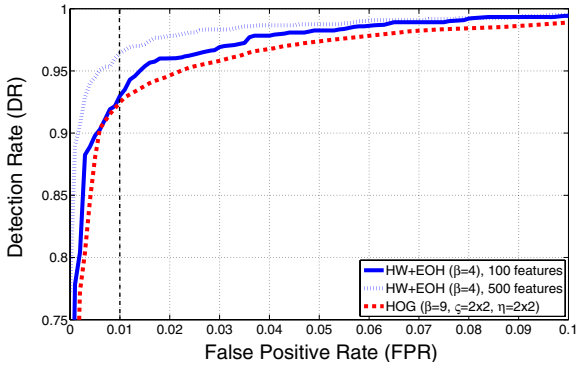

Fig. 6. Comparison between the HW+EOH based classifier and the best HOG based one

Table 1. Number of operations needed for HOG, HW and EOH

	+	×	/	$\sqrt{}$	>	*ii* accesses	built *ii*	edge mask	other
HOG	9,900	3,960	3,960	110	1,980	4,455	9	1D	bin interp.
HW	1,900	400	500	100	–	1,700	2	–	–
EOH	600	–	100	–	–	800	4	Sobel	bin interp.

Table 1 provides a summary of the basic operations needed by each feature set to perform the classification. In the case of HOG, 55 blocks are computed to classify each sample, spending in total all the indicated basic operations, building 9 integral images and computing a 1D edge mask for the sample. Next, we have detailed the operations needed supposing that all the features selected by Real AdaBoost are either HW (choosing filter (*c*) in Fig. 2, i.e., the slowest case) or EOH. In both cases, the classifier consists in 100 features, namely weak rules. As can be appreciated, the slowest HW+EOH classifier would consist in a 100 HW features, but it is still much faster than HOG.

5 Conclusions

This paper presents a comparison between two classifiers in an ADAS–oriented pedestrian database: Haar wavelets and edge orientation histograms (HW+EOH) features together with Real AdaBoost as learning algorithm, and the state–of–the–art best human–based classifier, histograms of oriented gradients (HOG) with SVM [5]. We describe the computation of the different features, and then tune their parameters to work with an ADAS pedestrian database. In this way, we provide a fair and accurated comparison in terms of detection rate and processing time (even comparing basic operations), which leads to conclude that HW+EOH based classifier achieves similar performance than HOG based one requiring much less processing time, concretely one order of magnitude faster.

Acknowledgments. This work was supported by the Spanish Ministry of Education and Science under project TRA2004-06702/AUT, BES-2005-8864 grant (first author) and Ramón y Cajal Program (fourth author).

References

1. Gavrila, D., Giebel, J., Munder, S.: Vision–based pedestrian detection: The PROTECTOR system. In: Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy (2004)
2. Shashua, A., Gdalyahu, Y., Hayun, G.: Pedestrian detection for driving assistance systems: Single–frame classification and system level performance. In: Proc. of the IEEE Intelligent Vehicles Symposium, Parma, Italy (2004)
3. Papageorgiou, C., Poggio, T.: A trainable system for object detection. IJCV 38(1), 15–33 (2000)
4. Levi, K., Weiss, Y.: Learning object detection from a small number of examples: the importance of good features. In: Proc. of the IEEE Conference on CVPR, Washington, DC, USA. pp. 53–60 (2004)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. of the IEEE Conference on CVPR. vol. 2., San Diego, CA, USA. pp. 886–893 (2005)
6. Schapire, R., Singer, Y.: Improved boosting algorithms using confidence–rated predictions. Machine Learning 37(3), 297–336 (1999)
7. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Conference on CVPR, Kauai, HI, USA (2001)
8. Lowe, D.: Distinctive image features from scale–invariant keypoints. IJCV 60(2), 91–110 (2004)
9. Zhu, Q., Avidan, S., Yeh, M.C., Cheng, K.T.: Fast human detection using a cascade of histrograms of oriented gradients. In: Proc. of the IEEE Conference on CVPR, New York, NY, USA (2006)