DETECTING PEOPLE IN IMAGES: AN EDGE DENSITY APPROACH

S. L. Phung, IEEE Member, and *A. Bouzerdoum,* IEEE Senior Member

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Wollongong, NSW 2522, Australia

ABSTRACT

In this paper, we present a new method for detecting visual objects in digital images and video. The novelty of the proposed method is that it differentiates objects from non-objects using image edge characteristics. Our approach is based on a fast object detection method developed by Viola and Jones. While Viola and Jones use Harr-like features, we propose a new image feature - the edge density - that can be computed more efficiently. When applied to the problem of detecting people and pedestrians in images, the new feature shows a very good discriminative capability compared to the Harr-like features.

Index Terms— people detection, image edge analysis, object detection, video surveillance, pattern recognition.

1. INTRODUCTION

Detecting people and pedestrians in images and video has applications in video surveillance, road safety and many others. For example, Collins et al. [1] developed a multi-camera surveillance system that can detect and track people over a wide area. Papageorgiou and Poggio [2] presented a vision system that is used in Daimler-Chrysler Urban Traffic Assistant to detect pedestrians. Haritaoglu and Flickner [3] described an intelligent billboard that uses a camera to detect and count the number of people in front of the billboard.

Existing methods for detecting people can be divided into two major categories. In the first category, people are detected using heuristic visual cues such as motion [4], background scene [5], silhouette shape [3] or color [6]. Image regions that may contain people can be identified rapidly by comparing a video frame with the previous frames or the background scene, or applying a color filter.

In the second category, pattern classifiers are trained to determine if each image window resembles the human body - a window is a fixed-size rectangular region of the image. This approach can cope well with image variations. Papageorgiou and Poggio [2] developed a pedestrian detection method, in which the Harr wavelet features are extracted from each 128-by-64 window and then classified using support vector machines. Recently, Viola and Jones [7] proposed a fast object detection method that relies on a cascade of classifiers. Each classifier uses one or more Harr-like features and is trained using the adaptive boosting (AdaBoost) algorithm. Their method has been applied successfully to the face detection problem.

This paper presents an object detection method that relies on object edge characteristics to differentiate objects and non-objects. We propose a new image feature called edge density that can be computed very fast. We apply the new method to detect people and pedestrians in images, and analyze the discriminative power of the edge density feature.

This paper is organized as follows. Section 2 describes the proposed object detection method and the new image feature. Section 3 focuses on an application of the proposed method in detecting people and pedestrians, and Section 4 is the conclusion.

2. EDGE DENSITY APPROACH

2.1. Overview

Our object detection method, which is motivated by Viola and Jones' system, scans exhaustively the windows of an input image. Because there could be over $200,000$ windows in a typical image of size 640×480 pixels, the classification method must be fast to support real-time detection. In our method, each window is processed by a cascade of strong classifiers to determine if it is an object or a non-object. If a strong classifier considers the window as a non-object, the window is immediately rejected; otherwise, the window is processed by the next strong classifier in the cascade. Because the majority of windows in an input image are non-object, the cascade structure reduces the average processing time per window.

A strong classifier is made up from one or more weak classifiers, and each weak classifier uses exactly one image feature extracted from the window. A weak classifier may have an error rate close to 0.5, but a strong classifier constructed using a boosting algorithm such as the AdaBoost [8] can have a lower error rate. The key idea of the AdaBoost al-

E-mails: phung@uow.eu.au and a.bouzerdoum@iee.org. This work is supported in part by research grants from the Australian Research Council and the University of Wollongong.

gorithm is to force each weak classifier to focus more on the training samples that the previous weak classifiers could not process correctly.

2.2. New Image Feature based on Edge Density

The system proposed by Viola and Jones [7] uses Harr-like features. A Harr-like feature is defined as the difference in the pixel sums of two adjacent regions. If a Harr-like feature is greater than a threshold, the weak classifier considers the window as an object. Essentially, a salient Harr-like feature indicates a window as an object if region A appears significantly darker or brighter than region B, where regions A and B are to be found through training. This strategy works well for objects with a defined inner structure such as the human face. For example, the eye region has a different brightness level compared to its surrounding. However, for some objects such as the human body (standing/walking pose) the dominant visual characteristics are the outer shape and edges. This observation motivates us to develop a new image feature that is based on edge density.

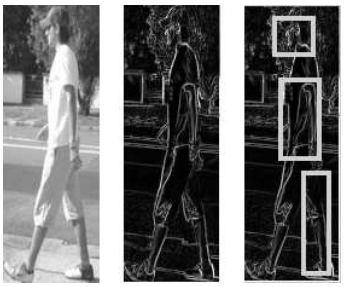


Fig. 1. *Left*: an image window. *Middle*: the edge magnitude. *Right*: three edge density features where each feature is the average edge magnitude in a specific subregion.

For a given window, an edge density feature measures the average edge magnitude in a subregion of the window (see Fig. 1). Let $i(x, y)$ be a window and $e(x, y)$ be the edge magnitude of the window. For a subregion r with the left-top corner at (x_1, y_1) and the right-bottom corner at (x_2, y_2), the edge density feature is defined as

$$f = \frac{1}{a_r} \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} e(x, y) \tag{1}$$

where a_r is the region area, $a_r = (x_2 - x_1 + 1)(y_2 - y_1 + 1)$.

If the edge density feature is greater (or smaller) than a threshold, the weak classifier considers the window as an object. This is equivalent to saying that a strong (or weak) presence of image edges in a subregion will determine if the window is an object. In a given window, there will be several thousands of subregions or features. The objective of system training is to identify the most salient features.

For the task of window scanning, there is a very efficient method to compute edge density features. Let $\mathbf{I} = \{I(x, y)\}$ be the input image of size $H \times W$. Let $\mathbf{E} = E(x, y)$ be its edge magnitude; $E(x, y)$ is found by applying edge operators, such as the difference, the Sobel or the Prewitt operators, on the entire image. The edge magnitude is a combination of the edge strength along the horizontal and vertical directions:

$$E(x, y) = \sqrt{E_h^2(x, y) + E_v^2(x, y)} \tag{2}$$

From the edge magnitude image $\mathbf{E}$, we compute an edge integral image $\mathbf{S}$. The pixel value $S(x, y)$ is defined as

$$S(x, y) = \sum_{x'=1}^{x} \sum_{y'=1}^{y} E(x', y') \tag{3}$$

That is, $S(x, y)$ is the sum of edge magnitudes in the rectangular region $\{(1, 1), (x, y)\}$.

Given the edge integral image, the edge density feature of a subregion $r = \{(x_1, y_1), (x_2, y_2)\}$ can be computed using only a few arithmetic operations:

$$f = \frac{1}{a_r}\{S(x_2, y_2) + S(x_1 - 1, y_1 - 1) - S(x_2, y_1 - 1) - S(x_1 - 1, y_2)\} \tag{4}$$

Our approach requires an extra computation for the edge magnitude image $\mathbf{E}$ before scanning occurs. Subsequently, each edge density feature involves only one subregion whereas each Harr-like feature involves at least two subregions. Hence, if the same number of features is used, the proposed approach can be expected to run fast. In Section 3, we shall study the classification performance of the new image feature.

2.3. Selecting the Most Salient Feature

A weak classifier is built by selecting the best feature from a feature pool of several thousands. This section describes the feature selection technique.

In a given training set, let w_1^+, w_2^+,...,w_M^+ be the weights of M training object patterns (i.e. positive patterns). Let w_1^-, w_2^-,...,w_N^- be the weights of N training non-object patterns (i.e. negative patterns). Let w^+ be the sum of all weights for object patterns, $w^+ = \sum_{i=1}^{M} w_i^+$. Let w^- be the sum of all weights for non-object patterns, $w^- = \sum_{i=1}^{N} w_i^-$. During training, we can modify individual weights [7], but the sum of w^+ and w^- is always equal to 1.

For a given feature f that corresponds to a subregion r, we first compute the cumulative histograms $c^+(\theta)$ and $c^-(\theta)$ for the object and non-object patterns, taking into account the pattern weights. There are two possible decision rules: (1) object if $f > \theta$, and non-object otherwise; (2) object if $f \leq \theta$, and non-object otherwise. Here, θ is a fixed threshold. The error rate for the first decision rule is

$$e_1(\theta) = w^- + c^+(\theta) - c^-(\theta) \tag{5}$$

The error rate for the second decision rule is

$$e_2(\theta) = w^+ - c^+(\theta) + c^-(\theta) \quad (6)$$

We select the decision rule that gives a smaller error. Hence, $e = min(e_1, e_2)$ is the error rate if feature f is used. From the feature pool, we choose the feature that gives the minimum error rate.

3. EXPERIMENTS AND ANALYSIS

We apply the proposed object detection approach to the problem of detecting people and pedestrians in images. In this section, we aim to study the process of building weak and strong classifiers for people detection, and the classification performance of the edge density feature.

3.1. Experiment Data

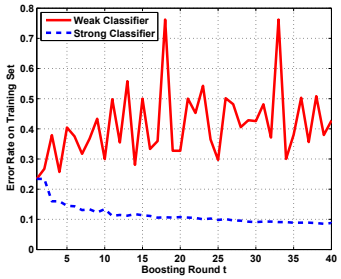
We collected a total of 2359 images that contain people and pedestrians, and manually identified the coordinates of the people in these images. From these images, 2664 people patterns were extracted. There are strong variations in the patterns: frontal view, side view, people in standing, bending, walking and running poses. We used 2000 patterns for training and 600 patterns for testing. In addition, from $10,000$ non-people images, we extracted 4000 patterns for training and 600 patterns for testing. Note that the training and test patterns were taken from disjoint sets of images. Examples of the people and non-people patterns are shown in Fig. 2.

Fig. 2. Examples of people and non-people patterns.

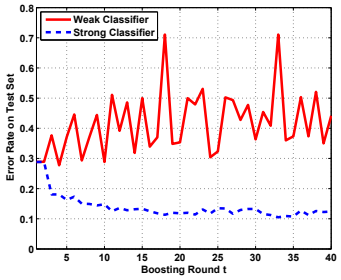
The aspect ratio (height/width) of the people patterns in our dataset has a mean value of 3.28 and a median value of 3.21. Note that the patterns include children as well as people in running or striding pose. Based on this result, we selected a window size of 60×18 pixels for designing the classifiers. This window size is found to reduce the computation load while keeping sufficient visual details for classification.

3.2. Analysis of Edge Density Features

The difference operators, $h_h = [1, -1]$ and $h_v = [1, -1]^T$, are used in the following experiments. We train a strong classifier for 40 rounds. In each round a weak classifier using exactly one edge density feature is formed. The weights of training patterns are modified according to the AdaBoost algorithm to put more emphasis on the patterns that the previous weak classifier incorrectly handles.



(a)



(b)

Fig. 3. Error rates of weak classifiers and a strong classifier on: a) the training set; b) the test set.

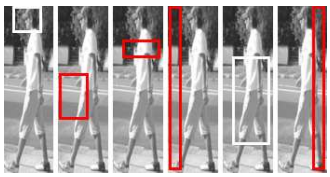


Fig. 4. Edge density features selected at boosting round $2^+, 14^-, 17^-, 32^-, 33^+$ and 37^-. The $+$ or $-$ superscript indicates if the feature uses decision rule (1) or (2).

Figure 3a shows the error rates of the strong classifier and weak classifiers as training progresses. The classification thresholds are set according to the AdaBoost algorithm [7]. The results show that the training error of the strong classifier decreases steadily with respect to the number of the training rounds. However, the error rates of individual weak classifiers fluctuate with an upward trend; this can be explained by the fact that each weak classifier focuses on more and more "dif-

ficult" patterns in the training set. After 40 training rounds, the strong classifier has an error rate of 11.3%.

Some edge density features selected by the strong classifier are shown in Fig. 4. These features indicate that the strong classifier mostly picks up the edge difference between the human body and the surrounding. For example, the feature selected at round 2 reflects the fact that there are strong edges in the human head region.

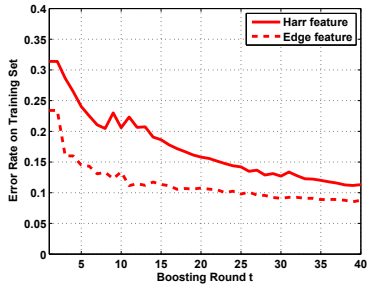
The performances of the strong classifier and individual weak classifiers on the test set are shown in Fig. 3b. The results show that even though the error rate of each weak classifier is high, the error rate of the strong classifier decreases steadily. In this case, there is little change in the error rate of the strong classifier after round 10. Using a validation set, we can detect when this occurs and stop training the strong classifier. At this point, we usually collect more data for training the next strong classifier and add it to the cascade.

The strong classifier using 10 features has an error rate of 14.8%, a false positive rate of 14.3%, and a false negative rate of 15.3%. For object detection purpose, we can use a cascade of strong classifiers, each of which is set to a low false negative rate (at a cost of a higher false positive rate).

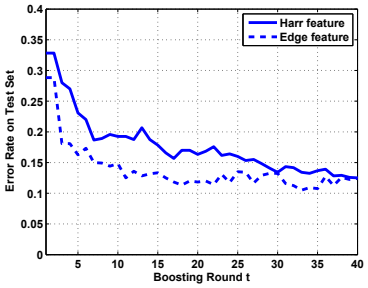
3.3. Comparison of Edge Density and Harr-like Features

For comparison purposes, we trained two strong classifiers: one using only edge density features, and the other using only Harr-like features [7]. The performances of the two strong classifiers on the training set and the test set are shown in Fig. 5.

The figure shows that the training error decreases faster using the edge density features. For example, after 10 rounds the training error is 13.4% for edge density feature, and 20.6% for Harr feature. The test error is also lower for the strong classifier that uses edge density features. For example, after 10 training rounds the best test error is 14.4% for edge density feature, and 18.7% for Harr feature. These results for the people detection task demonstrate a clear performance improvement using the proposed feature.



(a)



(b)

Fig. 5. Error rates of strong classifiers that use Harr-like and edge density features on: (a) the training set; (b) the test set.

4. CONCLUSION

We presented an object detection method based on a new image feature called edge density, which measures the presence or absence of image edges in a specific region of the object. The edge density feature can be computed very efficiently and it is found to have a better discriminative capability compared to the Harr-like features, when applied to the problem of detecting people in images.

5. REFERENCES

[1] R.T. Collins, A.J. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1456–1477, 2001.

[2] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection," in *Int. Conf. on Image Processing*, 1999, vol. 4, pp. 35–39.

[3] I. Haritaoglu and M. Flickner, "Attentive billboards," in *Int. Conf. on Image Analysis and Processing*, 2001, pp. 162–167.

[4] R. Patil, P.E. Rybski, T. Kanade, and M.M. Veloso, "People detection and tracking in high resolution panoramic video mosaic," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2004, vol. 2, pp. 1323–1328.

[5] S. Harasse, L. Bonnaud, and M. Desvignes, "Human model for people detection in dynamic scenes," in *Int. Conf. on Pattern Recognition*, 2006, vol. 1, pp. 335–354.

[6] Y. Rachlin, J. Dolan, and P. Khosla, "Learning to detect partially labeled people," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2003, vol. 2, pp. 1536–1541.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and application to boosting," *J. of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1995.