ROTATION INVARIANT FACE DETECTION USING SPECTRAL HISTOGRAMS AND SUPPORT VECTOR MACHINES

Christopher A. Waring and Xiuwen Liu

Department of Computer Science, Florida State University, Tallahassee, FL 32306

ABSTRACT

This paper presents a face detection method that detects faces with arbitrary rotation in the image plane. In this method, images are represented using a spectral histogram representation consisting of marginal distributions of filtered images. A support vector machine with an R.B.F. kernel is chosen as the classifier, which is trained on 4500 face and 8000 non-face images. The choice of filters allows a large degree of rotation invariance and by shuffling the marginals of certain filters, invariance to arbitrary rotation is achieved. A distinctive advantage of our method is that the invariance is achieved largely through the underlying representation while in other methods the invariance is typically achieved by detecting faces at a large number of different angles. The proposed method is tested on standard data sets and comparisons with other methods show that our method gives the best detection performance with respect to detection rate and false positives.

Index Terms— Feature extraction, object detection, image analysis, filtering, image processing, image classification, image recognition

1. INTRODUCTION

Face detection has received significant attention in recent years (see [6] for a recent survey). The problem of face detection from an image is commonly defined as follows [6]: Given an image $\mathbf{I}$, find all the occurrences of faces and the extent of each face in it. This formulation at core requires estimating the ideal face detection function g, which is defined as:

$$g(\mathbf{W}) = \begin{cases} 1 & \text{if } \mathbf{W} \text{ is a face image,} \\ 0 & \text{otherwise.} \end{cases}$$

Here $\mathbf{W}$ is the current region of interest. If the ideal $g(\mathbf{W})$ is available, one can detect all the faces by applying g to all possible $\mathbf{W}$'s in $\mathbf{I}$. A promising approach is to estimate $g(\mathbf{W})$ based on a large set of face and non-face images, collectively known as appearance-based methods, which can capture the complexity of the faces under a variety of conditions, and most recent methods belong to this category. Differences among appearance-based methods are centered around two main factors: the choice of the underlying representation and the estimation of $g(\mathbf{W})$. For a good representation, $g(\mathbf{W})$ would become simpler and thus a more accurate estimation is more likely. Given the choice of representation, one then needs to estimate $g(\mathbf{W})$ either directly using neural networks, support vector machines [3] or indirectly based on the Bayesian formulation by estimating the probability distributions of face and non-face images first.

In this paper, we propose an appearance-based method that uses a representation, called the *spectral histogram representation*, that is different from most existing methods. The spectral histogram representation captures the local structures of images through filtering and global structures implicitly through histograms of filtered images. Because of the properties of the spectral histogram representation, the proposed algorithm intrinsically provides a large degree rotation invariance. With its high detection performance, our system thus provides a counter example to the general belief that large invariances can not be achieved with accurate results (e.g. [2]).

The rest of this paper is organized as follows. Section 2 gives the spectral histogram representation and Section 3 describes the proposed face detection algorithm, including preprocessing, training, and post-processing stages. Section 4 shows the experimental results of the proposed method on face detection datasets and compares our results to that of others. Section 5 concludes the paper with a discussion.

2. SPECTRAL HISTOGRAM

The spectral histogram of an image is a feature vector consisting of the marginal distributions of filter responses of the image. Given an input image window $\mathbf{W}$, and a set of filters $\{F^{(\alpha)}, \alpha = 1, \ldots, K\}$, a sub-band image $\mathbf{W}^{(\alpha)}$ is computed through linear convolution given by $\mathbf{W}^{(\alpha)}(v) = F^{(\alpha)} * \mathbf{W}^{(\alpha)}(v) = \sum_u F(u)\mathbf{W}(v-u)$. For $\mathbf{W}^{(\alpha)}$, we define its histogram, a bin of which is given by

$$H_{\mathbf{W}}^{(\alpha)}(z_1, z_2) = \sum_{\vec{v} \in \mathbf{W}} \int_{z_1}^{z_2} \delta(z - \mathbf{W}^{(\alpha)}(v)) dz, \quad (1)$$

where z_1 and z_2 specify the range of the bin. The spectral histogram representation of $\mathbf{W}$ with respect to the chosen filters is defined as the concatenation of $\mathbf{H}_{\mathbf{W}^{(\alpha)}}$, given by $\mathbf{H}_{\mathbf{W}} = (\mathbf{H}_{\mathbf{W}^{(1)}}, \ldots, \mathbf{H}_{\mathbf{W}^{(K)}})$.

In this paper 33 filters are used, including 4 gradient filters, 5 Laplacian of Gaussian filters with different spatial scales, and 24 cosine Gabor filters with four different spatial scales and with six orientations at each scale at 0°, 30°, 60°, 90°, 120°, and 150°. One benefit of this choice of representation is the natural degree of rotation invariance. We are able to capture up to $\pm 45^\circ$ of rotation for the purpose of face detection. The spectral histogram of upright face images remains unchanged by the mirroring and the 180^0 rotation. This allows a system designed for upright face detection to be able to detect faces at $0^\circ \pm 45^\circ$ and $180 \pm 45^\circ$ degrees, for a total of 180 degrees of allowable rotation. Based on this large rotation invariance, we achieve the full rotation invariance by combining the results from the original image and from the image rotated by 90°. Because of the filters we used, image rotation can be achieved by shuffling the histograms. This result is straightforward as the response of a forward upright looking face to a Gabor filter oriented at 0° is the same as that from the same face rotated 90° to the corresponding Gabor filter at 90°.

2.1. Fast Implementation

Face detection requires evaluation of local spectral histograms of large number of different 21×21 windows and therefore fast computation is desirable. Note that calculating the histogram of local window $\mathbf{W}$ of a particular filter requires summation over all the pixels in $\mathbf{W}$. Let $\mathbf{W}$ be given by $(x_0, y_0) \leq \vec{u} \leq (x_1, y_1)$ and $\mathbf{I}^{(\alpha)}$ be the filtered image of image $\mathbf{I}$ using filter $F^{(\alpha)}$. By rewriting (1) in terms of $\mathbf{I}^{(\alpha)}$, we have

$$
\begin{aligned}
H^{(\alpha)}_{\mathbf{W}}(z_1, z_2) &= \textstyle\sum_{(x_0,y_0)\leq\vec{v}\leq(x_1,y_1)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz \\
&= \textstyle\sum_{(0,0)\leq\vec{v}\leq(x_1,y_1)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz \ + \\
&\textstyle\sum_{(0,0)\leq\vec{v}\leq(x_0-1,y_0-1)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz \ - \\
&\textstyle\sum_{(0,0)\leq\vec{v}\leq(x_0-1,y_1)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz \ - \\
&\textstyle\sum_{(0,0)\leq\vec{v}\leq(x_1,y_0-1)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz.
\end{aligned} \tag{2}
$$

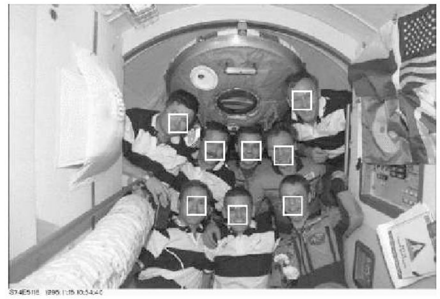
Here we assume $(0,0)$ is the top-left corner. If we let $\mathbf{W_0}$ be the rectangular window given by $(0,0) \leq \vec{u} \leq (x_1, y_1)$, $\mathbf{W_1}$ by $(0,0) \leq \vec{u} \leq (x_0 - 1, y_0 - 1)$, $\mathbf{W_2}$ by $(0,0) \leq \vec{u} \leq (x_0 - 1, y_1)$, and $\mathbf{W_3}$ by $(0,0) \leq \vec{u} \leq (x_1, y_0 - 1)$, we have $\mathbf{W} = \mathbf{W_0} + \mathbf{W_1} - \mathbf{W_2} - \mathbf{W_3}$. This gives the following

$$
\begin{aligned}
H^{(\alpha)}_{\mathbf{W}}(z_1, z_2) = H^{(\alpha)}_{\mathbf{W_0}}(z_1, z_2) + H^{(\alpha)}_{\mathbf{W_1}}(z_1, z_2) \\
- H^{(\alpha)}_{\mathbf{W_2}}(z_1, z_2) - H^{(\alpha)}_{\mathbf{W_3}}(z_1, z_2),
\end{aligned} \tag{3}
$$

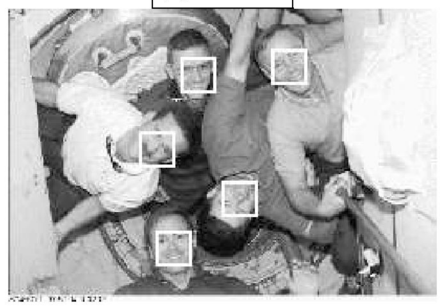
according to (2). Now if we define an *histogram integral image* $HI^{(\alpha)}_{(z_1,z_2)}$ for the bin, where the value at pixel (x, y) is $H^{(\alpha)}_{\mathbf{W}'}(z_1, z_2)$ and $\mathbf{W}'$ is the rectangle window given by $(0,0) \leq \vec{u} \leq (x, y)$. In other words, we define $HI^{(\alpha)}_{(z_1,z_2)}(x, y)$ as

$$
HI^{(\alpha)}_{(z_1,z_2)}(x, y) = \sum_{(0,0)\leq\vec{v}\leq(x,y)} \int_{z_1}^{z_2} \delta(z - \mathbf{I}^{(\alpha)}(\vec{v}))dz. \tag{4}
$$

Through the histogram integral image, as shown in (3), we can compute $H^{(\alpha)}_{\mathbf{W}}$ using L additions and $2 \times L$ subtractions, where L is the number of the bins in the histogram. An histogram integral image $HI^{(\alpha)}_{(z_1,z_2)}$ can be computed quickly using an algorithm given in [5]. Note that $\mathbf{I}^{(\alpha)}$ only needs to be computed once and field programmable gate array (FPGA) devices can be used for fast implementation.



(a) 8 / 8 / 0



(b) 5 / 5 / 0

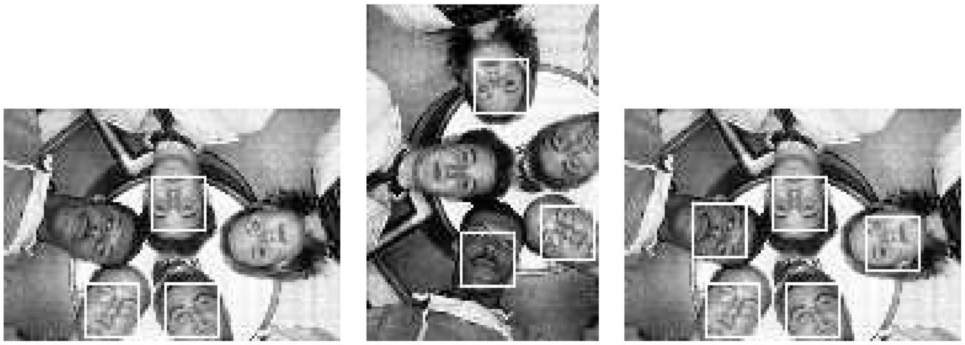
Fig. 2. Some typical Results on the test set. The first number corresponds to the number of faces in the image, the second corresponds to the number of faces found, and the third to the number of false detections.

3. TRAINING AND DETECTION

This section describes the preprocessing, training, and detection stages of the algorithm. Training is considered an appearance based case where 4500 face and 8000 non-face images are used. A support vector machine is used as the classifying function on a 21×21 image window.

3.1. Preprocessing and Training

The training data set used here is based on that provided by Henry Schneiderman from CMU. The data set consists



(a) (b) (c)

Fig. 1. An example of the face detection procedure. (a) The input image with upright detections. (b) The rotated input image with detections using the shuffled spectral histograms. (c) The final detection result by combining that in (a) and (b).

of 4500 randomly rotated face images and 8000 non-face images. A support vector machine (SVM) was selected as the classifying function. One of the main benefits of a large margin classifier is that it allows one to get an upper bound on the expected generalization error [3]. This is a distinctive benefit over traditional neural networks [4] as they seek to minimize the error on the training data and do not always generalize well. A support vector machine with an RBF (Radial Basis Function) kernel is trained on 4500 face and 8000 non-face images. The trained SVM is then used to classify windows in test images.

3.2. Detection

A three level Gaussian pyramid is built by successively down sampling the test image by a factor of 1.1. Each level of the pyramid is processed independently. This allows a degree of scale invariance as it allows the algorithm to search for faces at three different scales. For each image, a 21×21 image window is moved pixel by pixel through the entire image. The spectral histogram is then computed using the 33 chosen filters and stored. Once all the spectral histograms have been computed, they are fed to the trained SVM for classification. If a face is detected, then the coordinates of the center of the window are saved. The spectral histograms are then shuffled as described in Sec. 2 and once again fed to the SVM. If a face is detected, then the coordinates of the center window are also saved. A region growing algorithm is implemented to coalesce nearby detections in small regions. A thresholding is then applied to each region. Any region with fewer detections than the threshold is discounted as a non-face. After region growing and thresholding, regions that are less than three pixels apart are coalesced. The centroid for each new region is computed and saved. Once the entire image pyramid has been processed, detections at each layer are examined. A detection is marked as final if and only if it is found in at least two concurrent levels of the image pyramid. The results from the input image and the results from the shuffled histograms are then combined. A detection is registered as correct if it contains half or more of a face. Otherwise it is labeled a false detection.

Figure 1 shows an example. Fig. 1(a) shows the results using the upright face detection algorithm and Fig. 1(b) that based on the shuffled histograms. By combining the results, we achieve full rotation invariance as shown in Fig. 1(c).

4. EXPERIMENTAL RESULTS

The proposed algorithm was tested on the CMU rotated test set which is comprised of 50 images containing 225 faces. This test set as well as the face training set was provided by Henry Schneiderman from CMU. This test set has been used by Rowley et al. [4], Jeon et al. [1], and Jones and Viola [2]. Figure 2 shows another example on the test set with shows two images from the MIR space station. Fig. 2(b) is composed of faces at varying orientations which we are able to capture with no false positives. Fig. 3 shows an image with 135 faces all rotated at 45 degrees. Even though this image is the worst case for the proposed method (in that the detection at 45 degrees should be the worst compared to that at other angles), we are still able to detect 130 faces (about 96% of the faces in the image). Table 1 displays the results of our algorithm against the results of three other systems designed to be rotation invariant. Here the result of Jones and Viola's system [2] was estimated from their ROC curve. As this table shows ours is the one that performs the best with respect to false detections and detection rate.

5. CONCLUSION AND DISCUSSION

We have presented a method for rotation invariant face detection that performs the best on commonly used face detection datasets in terms of false positives and detection rate.

135 / 130 / 5

Fig. 3. Some typical results on the rotated test set. See Fig. 2 for the legend.

Table 1. Comparison on the rotated test set.

Method	Detection rate	False detections
Proposed method	93.0 %	137
Rowley et al. [4]	90.1%	303
Jeon, Lee, & Lee [1]	91.0%	196
Jones & Viola [2]	89.0%	137

We attribute the accurate detection results mainly to the desirable properties of the spectral histogram representation. By choosing proper filters for the spectral histogram representation, we achieve a large degree of rotation invariance with accurate detection results. To our best knowledge, ours is the first that achieves high detection accuracy with very large rotation invariance. This, to a large extend, invalidates the general belief that high detection accuracy is not feasible with large invariances (see e.g. [2]).

Besides the accurate detection results on both rotated and upright face detection datasets, another distinctive advantage of using the spectral histogram representation is the significant reduction of training data. For examples, Jones and Viola used over 100 million non-face training windows [2] while we only need 8,000 non-face training windows to achieve even better performance.

Acknowledgement

This work was supported in part by NSF grants CCF-0514743 and IIS- 0307998, and ARO grant W911NF-04-01-0268.

6. REFERENCES

[1] B. Jeon, S. Lee, and K. Lee, "Rotation invariant face detection using a model-based clustering algorithm," in *Proceedings of the IEEE International Conference on Multimedia and Exposition*, vol. 2, 2000, pp. 1149–1152.

[2] M. Jones and P. Viola, "Fast multi-view face detection," Technical Report TR2003-96, Mitsubishi Electric Research Laboratories, 2003.

[3] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proceedings of the IEEE Conference on Computer Vision and Patter Recognition*, 1997, pp. 130–136.

[4] H. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

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

[6] M. H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.