Robust Face Detection Using the Hausdorff Distance

Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz

BioID AG, Berlin, Germany
{o.jesorsky,k.kirchberg,r.frischholz}@bioid.com,
WWW home page: http://www.bioid.com

Abstract. The localization of human faces in digital images is a fundamental step in the process of face recognition. This paper presents a shape comparison approach to achieve fast, accurate face detection that is robust to changes in illumination and background. The proposed method is edge-based and works on grayscale still images. The *Hausdorff distance* is used as a similarity measure between a general face model and possible instances of the object within the image. The paper describes an efficient implementation, making this approach suitable for real-time applications. A two-step process that allows both coarse detection and exact localization of faces is presented. Experiments were performed on a large test set base and rated with a new validation measurement.

1 Introduction

Face recognition is a major area of research within biometric signal processing. Since most techniques (e.g. Eigenfaces) assume the face images normalized in terms of scale and rotation, their performance depends heavily upon the accuracy of the detected face position within the image. This makes face detection a crucial step in the process of face recognition.

Several face detection techniques have been proposed so far, including motion detection (e.g. eye blinks), skin color segmentation [5] and neural network based methods [3]. Motion based approaches are not applicable in systems that provide still images only. Skin tone detection does not perform equally well on different skin colors and is sensitive to changes in illumination.

In this paper we present a model-based approach that works on grayscale still images. It is based on the Hausdorff distance, which has been used for other visual recognition tasks [4]. Our method performs robust and accurate face detection and its efficiency makes it suitable for real-time applications.

2 Hausdorff Object Detection

The Hausdorff distance (HD) is a metric between two point sets. Since we want to use it for object detection in digital images, we restrict it to two dimensions.

2.1 Definition

Let $\mathcal{A} = \{a_1, \ldots, a_m\}$ and $\mathcal{B} = \{b_1, \ldots, b_n\}$ denote two finite point sets. Then the Hausdorff distance is defined as

$$H(\mathcal{A}, \mathcal{B}) = \max(h(\mathcal{A}, \mathcal{B}), h(\mathcal{B}, \mathcal{A})) \ , \quad \text{where} \tag{1}$$

$$h(\mathcal{A}, \mathcal{B}) = \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \|a - b\| \ . \tag{2}$$

Hereby $h(\mathcal{A}, \mathcal{B})$ is called the *directed Hausdorff distance* from set $\mathcal{A}$ to $\mathcal{B}$ with some underlying norm $|| \cdot ||$ on the points of $\mathcal{A}$ and $\mathcal{B}$.

For image processing applications it has proven useful to apply a slightly different measure, the (directed) *modified Hausdorff distance* (MHD), which was introduced by Dubuisson et al. [1]. It is defined as

$$h_{\text{mod}}(\mathcal{A}, \mathcal{B}) = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \|a - b\| \ . \tag{3}$$

By taking the average of the single point distances, this version decreases the impact of outliers making it more suitable for pattern recognition purposes.

2.2 Model-Based detection

Rucklidge [4] describes a method that uses the HD for detecting an object in a digital image. Let the two-dimensional point sets $\mathcal{A}$ and $\mathcal{B}$ denote representations of the image and the object. Hereby, each point of the set stands for a certain feature in the image, e.g. an edge point. The goal is to find the transformation parameters $p \in \mathcal{P}$ such that the HD between the transformed model $T_p(\mathcal{B})$ and $\mathcal{A}$ is minimized (see fig. 1). The choice of allowed transformations (e.g. scale and translation) and their parameter space $\mathcal{P}$ depends on the application. Efficient HD calculation allows an exhaustive search in a discretized transformation space.

The detection problem can be formulated as

$$d_{\hat{p}} = \min_{p \in \mathcal{P}} H(\mathcal{A}, T_p(\mathcal{B})) \ . \tag{4}$$

Then we call $h(T_p(\mathcal{B}), \mathcal{A}))$ the *forward distance* and $h(\mathcal{A}, T_p(\mathcal{B}))$ the *reverse distance*, respectively. To consider only that part of the image which is covered by the model, we replace the reverse distance by the *box-reverse distance* h_{box}. The definition can be found in [4].

3 System Description

The implemented face detection system basically consists of a coarse detection and a refinement phase, each containing a segmentation and a localization step. The following sections discuss these two phases in detail. Figure 2 gives a general overview of the described system.

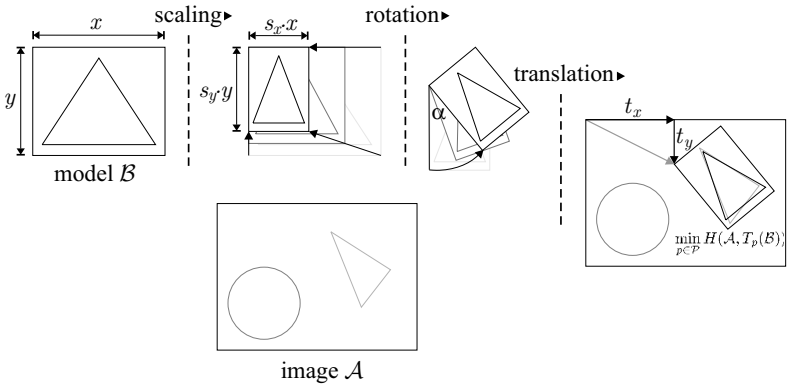


Fig. 1. Model fitting by scaling, translation and rotation.

Coarse Detection: Before applying any segmentation step, an area of interest (AOI) with preset width/height ratio is defined for each incoming image f. This AOI is resampled to a fixed size to be independent of the dimensions of f.

- *Segmentation:* An edge intensity image is calculated from the resized AOI with the Sobel operator. Afterwards, local thresholding guarantees that the resulting binary edge points are equally distributed over the whole image area.
- *Localization:* With a face model $\mathcal{B}$ and the binary representation $\mathcal{A}$ obtained by the segmentation step, a localization of the face in the image can now be performed according to equation (4). Experiments have proven that the modified forward distance is sufficient to give an initial guess for the best position. The parameter set $\hat{p}$ that minimizes $h(T_p(\mathcal{B}), \mathcal{A})$ is used as input for the refinement phase.

Refinement: Based on the parameter set $\hat{p}$, a second AOI is defined covering the expected area of the face. This area is resampled from the original image f resulting in a grayscale image h of the face area. Segmentation and localization are equivalent to the coarse detection step, except that a more detailed model $\mathcal{B}'$ of the eye region is used.

The values of the modified box reverse distance $h_{\text{box}}(\mathcal{A}', T_{\hat{p}'}(\mathcal{B}'))$ at the best position, especially when multiplied with the modified forward distance $h(T_{\hat{p}'}(\mathcal{B}'), \mathcal{A}')$, can be used to rate the quality of the estimation. This is helpful if a face/non-face decision is desired. The eye positions are calculated from the parameter sets $\hat{p}$ and $\hat{p}'$. Compared with manually set eye positions they are used to rate the quality of the system.

If the localization performance is not sufficient, an exact determination of the pupils can be achieved by additionally applying a multi-layer perceptron (MLP) trained with pupil centered images. The MLP localizer is not discussed in detail, but the results gained in combination with the rest of the system are included in the result section.

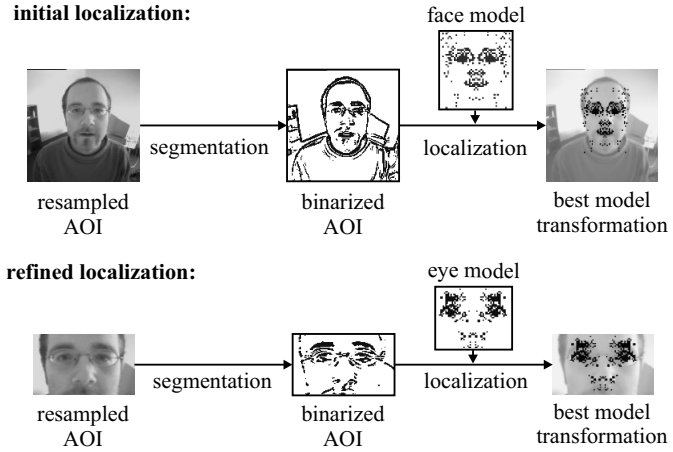


Fig. 2. The two phases of the face detection system containing the segmentation and the localization steps (AOI = area of interest). Top: coarse detection with a face model; bottom: refinement of the initially estimated position with an eye model.

Model Choice: The face and eye models, which are shown in figure 2, were initialized with average face data and optimized by genetic algorithms on a test set containing more than 10000 face images.

4 Validation

To validate the performance of our face detection system we introduce a relative error measure based on the distances between the expected and the estimated eye positions.

We use the maximum of the distances d_l and d_r between the true eye centers $C_l, C_r \in \mathbb{R}^2$ and the estimated positions $\widetilde{C}_l, \widetilde{C}_r \in \mathbb{R}^2$ as depicted in figure 3a. This distance is normalized by dividing it by the distance between the expected eye centers, making it independent of scale of the face in the image and image size:

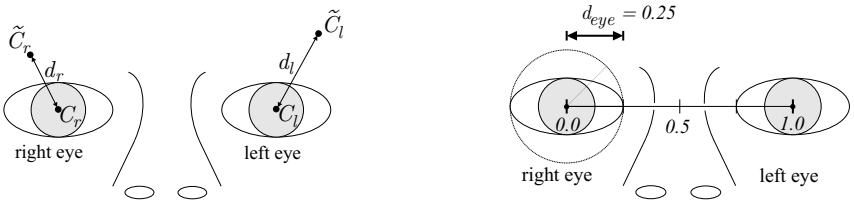
$$d_{eye} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|} \ . \tag{5}$$

In the following we will refer to this distance measure as *relative error*.

Considering the fact that in an average face the distance between the inner eye corners equals the width of a single eye, a relative error of $d_{eye} = 0.25$ equals a distance of half an eye width, as shown in figure 3b.

5 Experimental Results

Experiments on different test sets with different resolutions and different lighting and background conditions have been performed. The distribution function of the relative error between the expected eye positions (manually set) and the positions after each processing step has been calculated for each set. In this paper we present the results calculated on two test sets.



(a) (b)

Fig. 3. Relations and relative error. Figure (a) displays the relations between expected (C_l and C_r) and estimated eye positions ($\widetilde{C}_l$ and $\widetilde{C}_r$); (b) shows the relative error with respect to the right eye (left in image). A circle with a radius of 0.25 relative error is drawn around the eye center.

First one is the commonly used **extended M2VTS database** (XM2VTS) [2]. This database contains 1180 color images, each one showing the face of one out of 295 different test persons. Before applying the HD face finder we converted the images to grayscale and reduced their dimension to 360×288 pixel.

The second test set, which we will refer to as the **BIOID database**, consists of 1521 images (384×288 pixel, grayscale) of 23 different persons and has been recorded during several sessions at different places of our company headquarters. Compared to the XM2VTS this set features a larger variety of illumination, background and face size.

To give all researchers the opportunity to compare their results with ours, this test set is available for public at www.bioid.com/research/index.html (including manually set eye positions).

A comparison of some images of the two test sets can be seen in figure 4.

For our experiments we allowed translation and scaling. The search area was restricted to a square, horizontally centered region covering the whole image height. The scaling parameter range was chosen such that faces taking up between 25% and 55% of the image width could be detected. Figure 5 shows the

(a) (b) (c) (d) (e) (f)

Fig. 4. Image samples from XM2VTS (a, b) and the BIOID test set (c, d, e, f). Areas not considered during search have been brightened.

resulting distribution functions for both test sets. The results for HD search with additional MLP refinement have been included in the graphs.

If we consider a face found if $d_{eye} < 0.25$, the XM2VTS set yields 98.4% and the BIOID set 91.8% of the faces localized after refinement. This bound allows a maximum deviation of half an eye width between expected and estimated eye position (fig. 3b) as described in section 4.

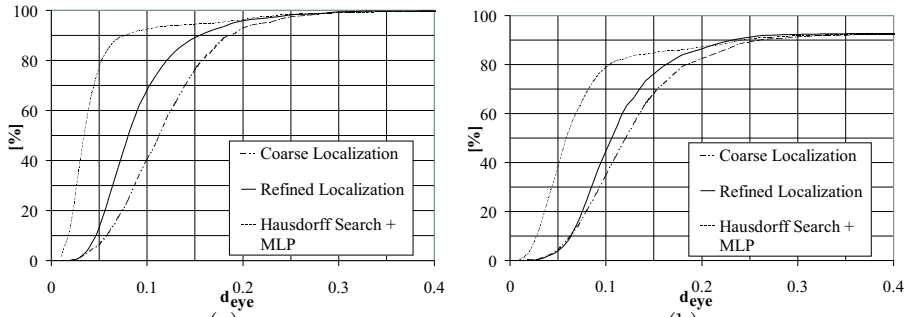


Fig. 5. Distribution function of relative eye distances for the XM2VTS (a) and the BIOID test set (b).

The average processing time per frame on a PIII 850 MHz PC system is 23.5 ms for the coarse detection step and an additional 7.0 ms for the refinement step, which allows the use in real time video applications (> 30 fps).

6 Conclusions and Future Research

In this paper we presented a face detection system that works with edge features of grayscale images and the modified Hausdorff distance. After a coarse detection of the facial region, face position parameters are refined in a second phase.

System performance has been examined on two large test sets by comparing eye positions estimated by the system against manually set ones with a relative error measure that is independent of both the dimension of the input images and the scale of the faces. The good localization results show that the system is robust against different background conditions and changing illumination. The runtime behaviour allows the use in realtime video applications.

Future research will concentrate on abolishing the restrictions of the detection of only frontal views and single faces, on automatic model creation and on transformation parameter optimization.

References

[1] M.P. Dubuisson and A.K. Jain. A modified Hausdorff distance for object matching. In *ICPR94*, pages A:566–568, Jerusalem, Israel, 1994.

[2] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, pages 72–77, March 1999.

[3] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 203–207, San Francisco, CA, 1996.

[4] W. Rucklidge. *Efficient Visual Recognition Using the Hausdorff Distance*, volume 1173 of *Lecture notes in computer science*. Springer, 1996.

[5] J. Terrillon, M. David, and S. Akamatsu. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In *Proc. of the Third International Conference on Automatic Face and Gesture Recognition*, pages 112–117, Nara, Japan, 1998.