Pattern Analysis & Applications (2000)3:9–18

Towards Detection of Glasses in Facial Images

X. Jiang[1], M. Binkert[1], B. Achermann[2] and H. Bunke[1]

[1]*Department of Computer Science, University of Bern, Bern, Switzerland;* [2]*Swisscom, Bern, Switzerland*

Abstract: In this paper we consider the automatic detection of the presence of glasses. Using the eye locations detected in facial images, two regions of interest around the eyes are investigated where glasses are expected. We introduce six measures for the likelihood of glasses. An experimental evaluation on two image sets demonstrates reasonable results of separating facial images with and without glasses. Furthermore, a combination of measures turns out to consistently improve the performance of the individual measures. We also perform a sensitivity analysis by means of simulation to figure out the inaccuracy of eye detection that can be tolerated by our glasses detection method.

Keywords: Facial images; Glasses detection; Salient features

1. INTRODUCTION

In dealing with large databases, indexing is an important aspect of object recognition in general, and face recognition in particular. To-date, face recognition experiments have been typically carried out on image databases of at most a few hundred persons. With the increasing sizes of image databases, the indexing problem will become more crucial. In this context, salient facial features may be useful, examples being gender, race, and approximate age (baby, child, young adult, senior adult).

The use of salient facial features requires the automatic detection of features for both automatic annotation of large image databases and indexing for fast face recognition. So far, however, the annotation of image databases has been widely done manually [1]. In fact, there exist only a few works in the literature describing the automatic detection of salient facial features. The task of gender determination was tackled by several authors [2–4]. O'Toole et al [4] explored the race classification problem. The first work in the image understanding community on age classification was reported by Kwon and Lobo [5]. In this paper, we consider another potential feature for image database indexing: the presence of glasses, which also possesses considerable discrimination power.

Our interest is the determination of the presence of glasses, but not a precise localisation of glasses. We are aware of only one earlier work [6] which considered the glasses detection problem. In Wiskott et al [6] a face is represented by a (possibly distorted) grid of feature points on the face. For a test image, this grid is generated by an elastic graph matching against a set of model grids. Then, facial parts of the test image relevant to a particular property (gender, glasses, beard, etc.) are compared to those of the models with respect to the similarity of local grey level distributions. The decision is finally done based on the annotation of the models.

In the present paper, we propose a new approach to glasses detection. In general, it is natural to expect a person to look straight into the camera, since also in human interaction people tend to turn their heads to look each other in the eyes. Accordingly, the main interest in this work is front view images, i.e. facial images with a person looking straight into the camera. Our approach to the detection of glasses starts with the localisation of the eyes. Given the eye locations, we define facial regions where glasses are expected, and measures for evaluating the presence of glasses. We also suggest a combination of measures, and experimentally demonstrate the superiority of the combination measures. Moreover, a simulation is carried out to investigate the sensitivity of our approach with respect to

Received: 14 December 1998
Received in revised form: 30 June 1999
Accepted: 8 July 1999

the inaccuracy of eye location detection. Our method is tested on an image set acquired in our lab, and on the Olivetti Research Laboratory (ORL) image set [7].

In the remainder of the paper, we first describe the eye detection method used in our experiments. Then, we propose measures of the presence of glasses in Section 3 and the combination of measures in Section 4. The sensitivity analysis of the measures w.r.t. the eye detection inaccuracy is given in Section 5. A short discussion of non-frontal views is presented in Section 6. While the experimental results on our image set are discussed in Sections 3 and 4, those for the ORL image set are summarised in Section 7. Finally, some discussions conclude the paper.

2. DETECTION OF EYE LOCATION

The detection of facial features, in particular the eyes, has many applications. Measurements of eye separation, for instance, have been used to provide invariance to distance from the camera and to normalise other face measurements [8,9]. In our case, the eye location defines the facial region in which we search for clues as to the presence of glasses. In the literature, a number of methods have been proposed to solve the eye detection problem. Some examples are Hough transform-based approaches [10–12], projection analysis [9], knowledge-based search [13], structure analysis of preattentive features [14], and deformable models [15–17]. A detailed discussion on facial feature detection can be found in Chellappa et al [18].

For the purpose of glasses detection, any eye localisation method can be used. To experimentally evaluate our glasses detection approach we have adopted the Hough transform approach introduced by Kothari and Mitchell [11], with some modifications. For reasons of completeness, we give a brief description of the method and our modifications.

Here the assumption is made that the face is not strongly rotated, i.e. the eyes are at similar heights in facial images. Generally, the most significant feature of eyes is the iris' circular shape and the strong contrast between the iris and the sclera. In particular, the lines drawn at edge points along the gradient direction (from light to dark) must intersect at the iris' centre. The detection of this centre is based on an accumulation array. For all edge points with a significant gradient magnitude, the line along the gradient direction is considered, and all cells in the accumulation array corresponding to points on this line are incremented. Then, a set of the largest accumulation values are selected as eye candidates, some of which may be caused by patterns on other parts of the face, the subject's clothes, or the background. All pairs of eye candidates are evaluated by a cost function. Finally, the pair with the smallest cost is chosen as the eyes.

Basically, we have made two modifications. First, the computation of gradients and lines along the gradients is replaced by a simpler strategy. For each pixel we consider the pixel in the 3×3 neighbourhood with the smallest grey level. If this minimum is significantly smaller than the current pixel, then a line is drawn from the current pixel to the minimum pixel, and the accumulation cells on this line are incremented. Now there are merely four predefined line directions (horizontal, vertical and diagonal), and the determination of involved accumulation cells becomes trivial. The second modification concerns the cost function for evaluating pairs of eye candidates. Given a pair of eye candidates at location p_{ij} and $p_{i'j'}$ with accumulation values H_{ij} and $H_{i'j'}$, respectively, the cost function in Kothari and Mitchell [11] is defined by

$$C = (H_{ij} - H_{i'j'})^2 + \lambda(i - i')^2$$

which is to be minimised. It is necessary that the two candidates have similar accumulation values and are at similar heights in the image. The coefficient λ controls the relative weighting of the two criteria. In our implementation we have incorporated two additional criteria into the cost function. The accumulation values H_{ij} and $H_{i'j'}$ themselves should be high, otherwise a pair of eye candidates with similar but low accumulation values would receive a favourable evaluation. Moreover, the neighbours of the eye candidates should also have high accumulation values. Considering all four criteria, we get a new cost function:

$$C^* = (H_{ij} - H_{i'j'})^2 + \lambda_1(i - i')^2 - \lambda_2(H_{ij} + H_{i'j'}) - \lambda_3 \sum_{p_{kl} \in N(p_{ij}) \cup N(p_{i'j'})} H_{kl}$$

where $N(p_{ij})$ denotes the set of eight direct neighbours of p_{ij}. In our experiments, we have set $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 1$. Our experience is that the modified version of the eye detection method achieves a slightly better performance at a higher speed.

As a final remark, we want to point out the potential of this eye localisation method in dealing with complex background. In our test images, the test person stands in front of a uniformly bright background (see Section 3). Nevertheless, a complex non-uniform background can be tolerated as long as there are no grey level patterns in the images that are similar to eyes, and which can therefore potentially confuse the eye localisation algorithm.

3. MEASURING THE PRESENCE OF GLASSES

Given the approximate location of the eyes, in this section we propose measures for deciding on the presence of glasses. The glasses are characterised by high grey level discontinuities of the frame against the facial part. The glasses' frame is usually visible around the eyes (above, below and between the eyes). In the area above the eyes, however, it is difficult to distinguish between grey level discontinuities caused by glasses and those caused by the eyebrows. The regions below and between the eyes which remain are generally free of any foreign influence. Certainly, the region below the eyes must be carefully dimensioned, otherwise the nose would fall into this region. We assume that in these two areas the majority of (high) grey level discontinuities are caused by the glasses' frame and its shadow on the face.

Therefore, the edgeness measure of these regions, referred to as frame regions F_1 and F_2 in the following discussion, provides a clue as to the presence of glasses.

Figure 1 shows the frame regions specified in terms of the distance d between the eyes and four parameters p, q, r and s. The frame region F_1 is defined as a rectangular area below the eyes of a length pd and a height qd. The upper side of F_1 is located at a distance rd from the middle point A between the eyes. Finally, F_1 is horizontally centred at A. The frame region F_2 is assumed to contain the bridge connecting the two glasses parts. The presence of a bridge (i.e. glasses) should cause higher edgeness values than an absence. As long as F_2 is not chosen to be too large such that undesired influences from the eyes are introduced, it also provides useful clues as to the presence of glasses. We define F_2 as a rectangular area of height $2rd$ and width sd centred at the middle point A between the eyes (see Fig. 1).

Within the frame regions, clues as to grey level discontinuities are gathered. Note that we are only interested in the presence and magnitude of grey level discontinuities, not in their precise localisation. Therefore, we apply a simple edgeness measure based on the sum of the absolute value of derivatives in the x and y direction:

$$\text{edgeness}_{ij} = |g_{i+1,j} - g_{i-1,j}| + |g_{i,j+1} - g_{i,j-1}|$$

where g_{ij} represents the grey level at pixel p_{ij}. We try to expand the edgeness response of the frame border to the inner pixels of the frame so that not only the frame border, but more pixels of the frame, contribute to the accumulation of edgeness clues. This is done by iterating the edge detector k times: each time, the result of the last iteration is used as the input to the next iteration. In our current implementation, we have chosen $k = 2$. After this operation we perform a grey scale dilation with a 3×3 flat structuring element, i.e. the edgeness value of each pixel is replaced by the maximum in its 3×3 local neighbourhood. In this way, the high edgeness values can be propagated into their neighbourhood. Again, the grey scale dilation can be iterated to produce a final edgeness map. Currently, we perform three iterations.

After computation of the edge measures, a simple clue as to the presence of glasses is the average edgeness measure of all pixels in the frame region F_1:

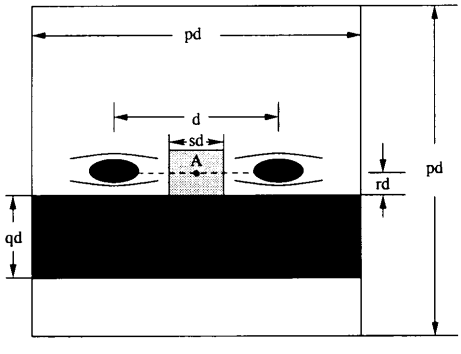

Fig. 1. Frame region F_1 (black region) and F_2 (grey region), and the region for automatic threshold determination.

$$M_1 = \left[\sum_{p_{ij} \in F_1} \text{edgeness}_{ij} \right] / |F_1|$$

where $|F_1|$ represents the number of pixels in the frame region. Alternatively, we can perform a thresholding operation to generate a binary edge map of F_1 and count the number of edge points. There exist a large number of methods for automatic determination of thresholds [19,20]. In our current implementation, we have used the discriminant analysis technique proposed by Otsu [21]. The threshold determination is not applied to the frame region F_1. The reason is that when there are no glasses, there are only a few edge points in F_1, and the computed threshold is not useful. Instead, we consider a quadratic region of side length pd centred at the middle A of the eyes; see the larger unshaded region in Fig. 1. For this region, the edge detection and the maximum operation are applied. The histogram of the edge detection results is the input to Otsu's algorithm for threshold determination. This yields a second clue as to the presence of glasses:

$$M_2 = \left[\sum_{p_{ij} \in F_1} \textbf{if } \text{edgeness}_{ij} \geq T \textbf{ then } 1 \textbf{ else } 0 \right] / |F_1|$$

where T is the determined threshold. This represents the percentage of edge points within F_1. In the measure M_2 all edge points contribute equally to the clue accumulation. In this way, however, weak edge points in the facial part of the frame region may provide unnecessarily high support, and thus bias the clue accumulation. Alternatively, we can take the magnitude of edge points into account and formulate a third clue as to the presence of glasses:

$$M_3 = \left[\sum_{p_{ij} \in F_1} \textbf{if } \text{edgeness}_{ij} \geq T \textbf{ then } \text{edgeness}_{ij} \textbf{ else } 0 \right] / |F_1|$$

In a similar way, the evidence of glasses for the second frame region F_2 can be quantified in terms of

$$M_4 = \left[\sum_{p_{ij} \in F_2} \text{edgeness}_{ij} \right] / |F_2|$$

$$M_5 = \left[\sum_{p_{ij} \in F_2} \textbf{if } \text{edgeness}_{ij} \geq T \textbf{ then } 1 \textbf{ else } 0 \right] / |F_2|$$

$$M_6 = \left[\sum_{p_{ij} \in F_2} \textbf{if } \text{edgeness}_{ij} \geq T \textbf{ then } \text{edgeness}_{ij} \textbf{ else } 0 \right] / |F_2|$$

Note that the same threshold T determined for M_2 and M_3 is applied to M_5 and M_6. In general, we expect that the presence of glasses will give abundant support to all six measures defined above. Thus, facial images with and without glasses should result in a bimodal distribution, making a separation of the two kinds of facial images possible.

We demonstrate the usefulness of our measures using an image set acquired in our lab which consists of 30 persons (mainly Europeans and a few Asians), and which has been used for testing face recognition systems [22,23]. Acciden-

tally, exactly half of the people wear glasses. For each person, 10 grey level images are available with varying view directions (two frontal views, two looking to the left, two looking to the right, two downwards and two upwards). All images are of the same resolution 512×342, and were taken under normal lighting conditions. Since our main interest in this work concerns facial images with frontal views, we first look at the results on this subset S_f of 60 images. The results of other view directions will be discussed in Section 6. In our experiments we have fixed the parameters $p = 2$, $q = 0.5$, $r = 0.17$ and $s = 0.45$ for the frame regions (the parameterisation aspect is discussed later in this section). Figure 2 visualises the edgeness measures for three people from the image set S_f. The eyes detected are depicted by dark crosses. The two frame regions (cf. Fig. 1) are drawn as well. In all 60 images, the eye position has been precisely localised. Since the size of the evaluation regions is dependent on the distance between the two eyes, our approach is able to adapt itself to different head sizes in the images. The head size of the person in Fig. 2(c), for instance, is quite different from that in Figs 2(a) and (b). The edgeness measures of the regions F_1 and F_2 visualised in Fig. 2 confirm the expectation that the majority of significant edgeness values are caused by the glasses. On the entire image set S_f, the values of the six measures are histogrammed in Fig. 3, where samples with and without glasses are denoted by the plus and circle symbol, respectively. The two classes of facial images are well separated. Intuitively, Fig. 3 reveals a superior class separation of the measures M_4, M_5 and M_6 compared to M_1, M_2 and M_3, respectively. A quantitative characterisation of class separation is Fisher's criterion [24]:

$$J(D_1,D_2) = \frac{(\hat{m}_1 - \hat{m}_2)^2}{\sigma_1^2 + \sigma_2^2}$$

where $\hat{m}_1$ and $\hat{m}_2$ represent the mean of two distributions D_1 and D_2, respectively. The terms σ_1^2 and σ_2^2 correspond to the variance of D_1 and D_2, respectively, indicating the within-class scatter of the distributions. Using Fisher's criterion, the goodness of the six measures M_k on the test data in terms of class separation is quantitatively characterised by:

	M_1	M_2	M_3	M_4	M_5	M_6
J	8.12	12.69	7.83	16.25	37.02	19.96

confirming the superiority of M_4, M_5 and M_6 over their respective counterpart.

To compute the measures of the presence of glasses, we first have to fix the dimensional parameters of the frame regions as defined in Fig. 1. For this purpose we have considered three potentially feasible values of each parameter: $p = 2 \pm 0.2$, $q = \frac{1}{2} \pm 0.1$, $r = \frac{1}{7} \pm 0.03$ and $s = \frac{1}{3} \pm 0.12$, resulting in a total of 81 candidate configurations. The goodness of each candidate configuration is evaluated by computing Fisher's criterion J for all six measures M_k. The optimal one has been chosen so that the evaluation function

$$\max(J(M_1), J(M_2), J(M_3)) + max(J(M_4), J(M_5), J(M_6))$$

is maximised, which turned out to be $p = 2$, $q = 0.5$, $r = \frac{1}{7} + 0.03 \approx 0.17$ and $s = \frac{1}{3} + 0.12 \approx 0.45$. This particular parameterisation has been used in our experiments, and generated the results reported in this section.

4. COMBINATION OF MEASURES

So far the two frame regions F_1 and F_2 have been investigated separately, and the experimental results suggest the superiority of the measures of F_2 over their counterparts of F_1. Both frame regions provide valuable information about the presence of glasses. We believe that a combination of both information sources enables an even more reliable decision.

To verify this, we have investigated linear combinations:

$$M_{kl} = M_k + \beta M_l, \quad k = 1,2,3;\ l = 4,5,6 \tag{1}$$

Each combination measure M_{kl} is optimised by considering

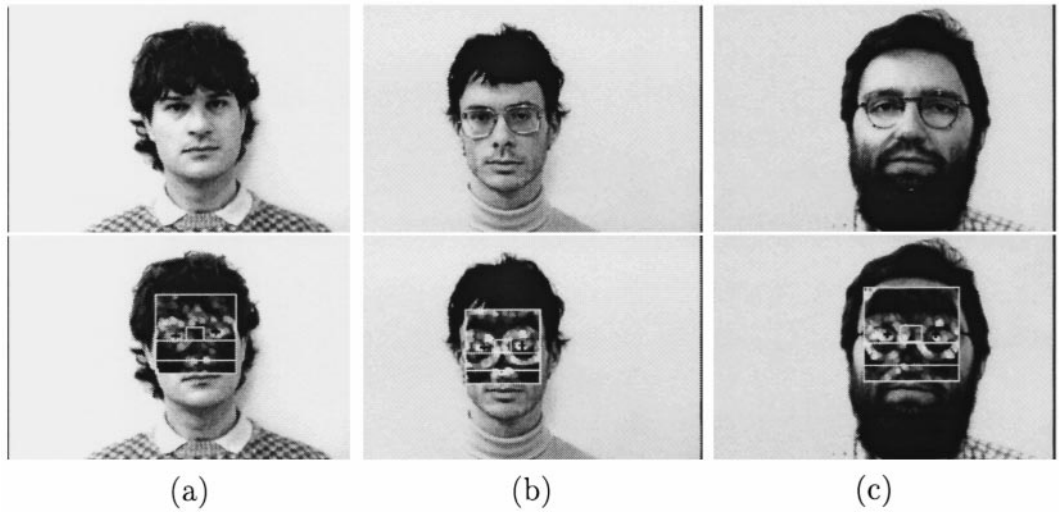

Fig. 2. Three facial images with the detected eyes and the edgeness values.

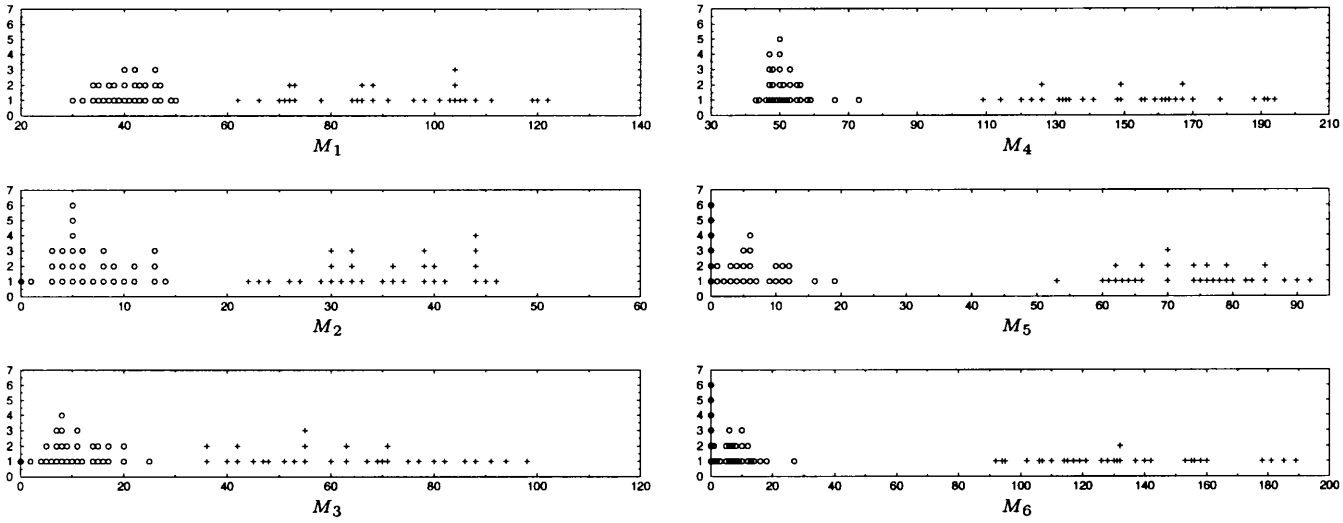

Fig. 3. Histograms of the measures.

40 β values from 0 to 3.9 at a step of 0.1. For each value of β we compute the measure M_{kl} for the image set S_f. Fisher's criterion J provides us with the class separability for every β value, and the optimal β value is chosen to be the one that gives the best J rating. The optimal measures M_{kl} are listed in Table 1, containing the corresponding β value, Fisher's criterion J, and the distribution of the two classes. Regarding the later case, D_1 and D_2 represent the class of images without and with glasses, respectively, and the interval of their measures.

In Table 1, the relationship

$$J(M_{kl}) > J(M_k) \quad \text{and } J(M_{kl}) > J(M_l) \tag{2}$$

holds, i.e. the combination measure M_{kl} consistently improves the class separability compared to the individual measures M_k and M_l. It is worth mentioning that Fisher's criterion J is quite stable around the optimal β value. With respect to the J value, M_{25} turns out to be the overall best combination measure in this experiment. The corresponding β may be chosen from the interval [1.3,2.3]; for all values from that interval $J \geq 42$ holds.

Note that in linear combinations (1) we have kept the coefficient of M_k as 1. The reason for this restriction lies simply in the fact that from the class separability point of view (in terms of Fisher's criterion J), a linear combination $aM_k + bM_l$ is equivalent to $M_k + \frac{b}{a}M_l$, which is covered in the investigated linear combinations (1).

5. SENSITIVITY ANALYSIS

A fundamental assumption of our approach to glasses detection is eye localisation with a 'sufficient' accuracy. Under this condition it is expected that the two facial image classes will be well separated by the measures proposed. Now we are interested in a quantitative characterisation of sufficient eye localisation accuracy, i.e. we ask what is the maximal eye localisation error that can be tolerated by the measures without affecting the class separability.

For this purpose, we have carried out a series of simulations on the image set S_f described in Section 3. We manually marked the centre of the iris for all images. This point is regarded as the ideal eye position. Then, we consider the pixels P_n with a chessboard distance $3n$, $n = 0, 1, 2$, from the iris' centre as instances of erroneous eye positions. In the nth ($n = 0, 1, 2$) simulation, we assume that the eye localisation accuracy is within $3n$ pixels from the ideal position. For $n = 0$, it means that only the (manually specified) ideal eye position P_0 is allowed. On the other hand, P_0 and all P_1s are potential eye positions from the eye localisation for $n = 1$. Similarly, at an inaccuracy of 6 pixels ($n = 2$), the eye position may be any of P_0, P_1s and P_2s. In our image set S_f the iris' contour is typically bounded by P_2's. See Fig. 4 for an illustration of the simulated eye positions and the simulation area in case of a particular image (that in Fig. 2(a)). In each simulation we independently choose one of the allowable eye positions for each

Table 1. Optimised combination measures

	M_{14}	M_{15}	M_{16}	M_{24}	M_{25}	M_{26}	M_{34}	M_{35}	M_{36}
β	3.5	3.8	3.1	0.4	2.1	0.5	3.0	3.8	2.6
J	16.58	39.76	20.58	21.24	42.44	24.91	16.76	38.55	20.68
D_1	[188,305]	[30,122]	[30,133]	[19,43]	[1,53]	[1,27]	[140,243]	[2,97]	[2,95]
D_2	[453,800]	[297,463]	[357,707]	[75,122]	[148,234]	[73,139]	[373,675]	[269,435]	[286,585]

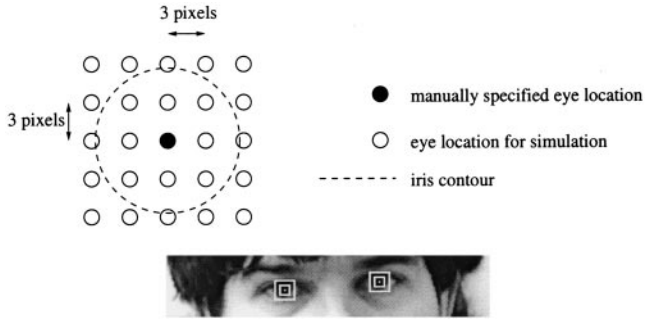

Fig. 4. Simulated eye positions.

eye. For all possible pairs of eye positions, the six measures M_k are computed for all facial images in S_f. We have, therefore, $60 \cdot (2n + 1)^4$ test instances in total for the nth simulation. The results of the simulation series are shown in Table 2, where the 'Overlap' row describes the percentage of test instances for which the measure lies in the overlapping interval of the two classes. Obviously, the class separation in terms of J decreases with increasing inaccuracy of eye localisation. In the case of ideal eye localisation, the two classes are well separated without overlap. At an inaccuracy of 3 pixels ($n = 1$), the class separation remains good: there is a small overlapping value domain for M_1, M_2 and M_3, while M_4, M_5 and M_6 are free of overlap. For simulation 2 with an inaccuracy of 6 pixels, the overlap becomes larger, particularly for M_1, M_2 and M_3.

Improved class separation can be achieved by the combination of measures as discussed in the last section. For simulation 2, the optimal combination results are listed in Table 3. Again, relationship (2) holds, implying that the combination measure M_{kl} consistently improves the class separability compared to the individual measures M_k and M_l. Particularly interesting is the fact that even in simulation 2 with an eye localisation inaccuracy of 6 pixels, all combination measures now become free of overlap.

The founding of this simulation can be summarised as follows. The measures M_4, M_5 and M_6 defined for frame region F_2 generally enable a better class separation than those for F_1, as already observed in the experiments reported in Section 3. Moreover, an edge localisation accuracy within the iris' contour is sufficient for a good class separation. It is important to notice that this accuracy requirement seems to be achievable by the eye detection techniques known in the literature. In Benn et al [10] and Huang et al [14], for instance, a localisation inaccuracy of less than three pixels has been reported. This fact manifests the practical applicability of the approach proposed in this paper.

6. INVESTIGATION OF NON-FRONTAL VIEWS

Our main interest in this work concerns facial images with frontal views. For this reason, the experimental efforts have been concentrated on the image set S_f. For completeness, we have also evaluated the behavior of our measures on the 240 images of our test set that have non-frontal view directions. In nine images the eye localisation method described in Section 2 made a false decision, and we have therefore excluded them from further consideration. For our evaluation, the 231 remaining images were divided into

Table 2. Simulation results

	M_1	M_2	M_3	M_4	M_5	M_6
			Simulation 0			
J	7.90	13.13	7.48	15.41	30.88	17.55
D_1	[32,56]	[5,14]	[5,26]	[44,67]	[0,17]	[0,22]
D_2	[65,109]	[24,44]	[37,86]	[106,197]	[54,93]	[79,194]
Overlap (%)	0.00	0.00	0.00	0.00	0.00	0.00
			Simulation 1			
J	7.56	11.40	7.09	15.02	31.07	17.31
D_1	[30,67]	[0,20]	[1,35]	[40,71]	[0,22]	[0,25]
D_2	[59,128]	[20,51]	[32,108]	[97,201]	[43,94]	[66,195]
Overlap (%)	4.47	0.04	0.93	0.00	0.00	0.00
			Simulation 2			
J	6.52	8.85	6.09	13.09	24.77	14.74
D_1	[28,77]	[0,31]	[0,45]	[34,76]	[0,34]	[0,39]
D_2	[54,146]	[17,59]	[26,119]	[82,209]	[30,96]	[48,206]
Overlap (%)	19.27	20.02	14.26	0.00	0.06	0.00

Table 3. Optimised combination measures for simulation 2

	M_{14}	M_{15}	M_{16}	M_{24}	M_{25}	M_{26}	M_{34}	M_{35}	M_{36}
β	2.1	3.7	1.6	0.5	1.6	0.4	2.1	3.9	1.6
J	13.88	27.75	16.10	17.43	28.68	19.29	14.00	26.94	16.09
D_1	[118,218]	[29,177]	[29,114]	[21,59]	[0,76]	[0,39]	[79,189]	[0,158]	[0,88]
D_2	[228,562]	[215,486]	[153,450]	[67,154]	[85,205]	[50,131]	[204,538]	[192,480]	[129,426]

Fig. 5. Examples of non-frontal views.

subsets S_d (downwards), S_u (upperwards) and S_s (sideviews). Because of the symmetry between looking to the left and right, these two subgroups were put together into the subset S_s. A representative from each of the three subsets is shown in Fig. 5, and the six measures M_k on each subset are histogrammed in Figs 6 – 8. For this experiment, we have not tuned the configuration parameters of the frame regions; instead, the optimal values determined for the frontal view have been used.

For the downward view, the measures M_1, M_2 and M_3 reveal a similar distribution, and thus class separability, as in the case of frontal images S_f. On the other hand, the class separability w.r.t. M_4, M_5 and M_6 become substantially worse. In this view direction the frame region F_2 more likely

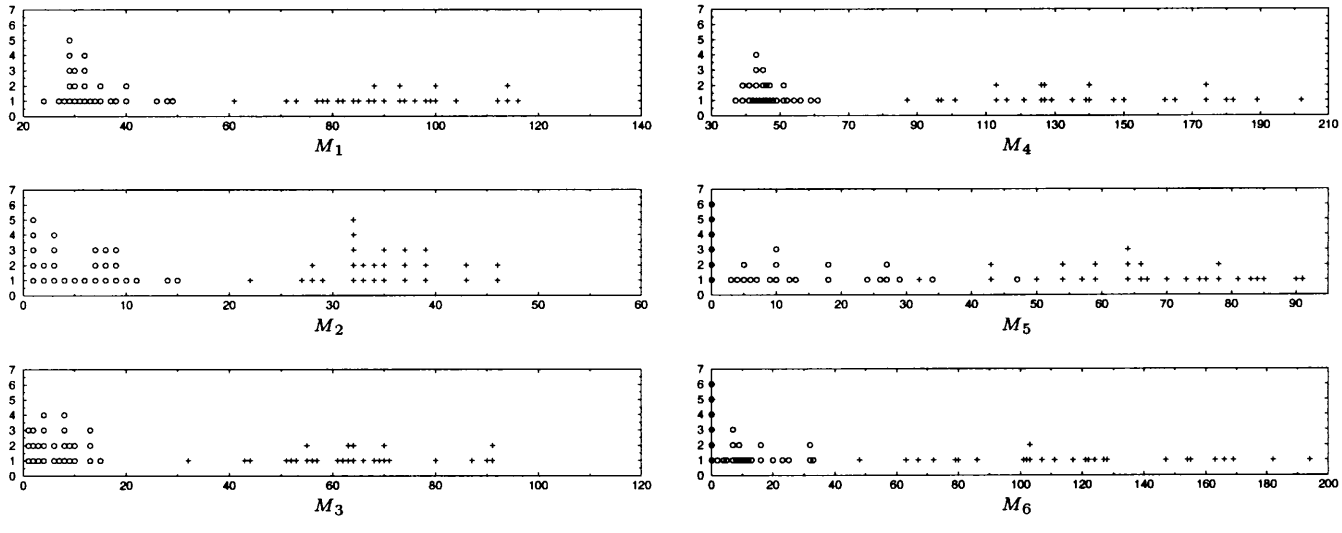

Fig. 6. Histograms of the measures M_k for the image subset S_d.

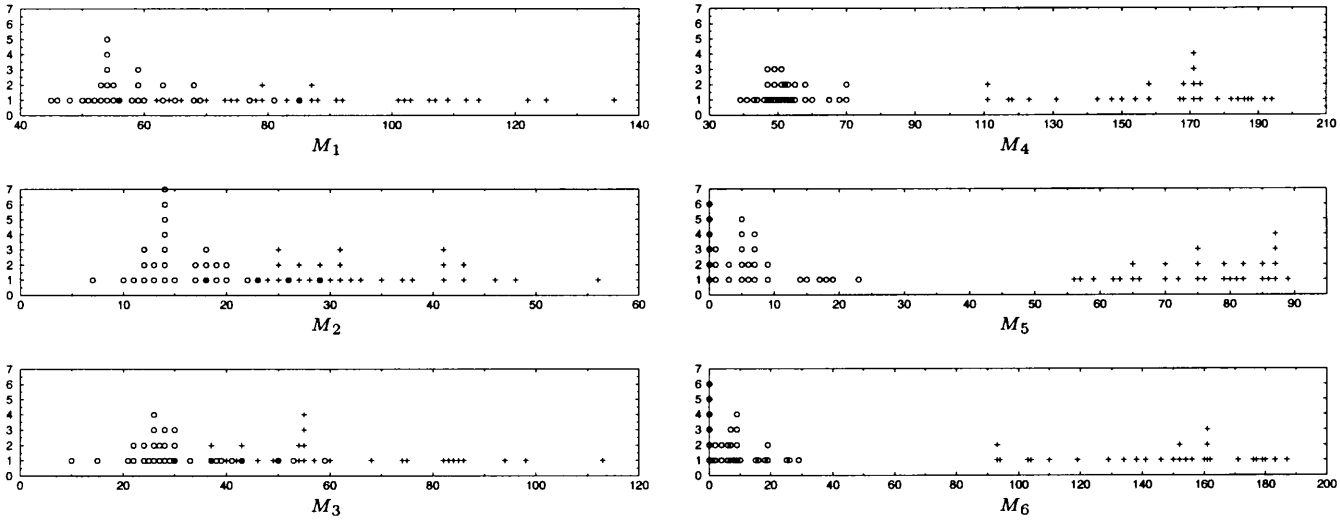

Fig. 7. Histograms of the measures M_k for the image subset S_u.

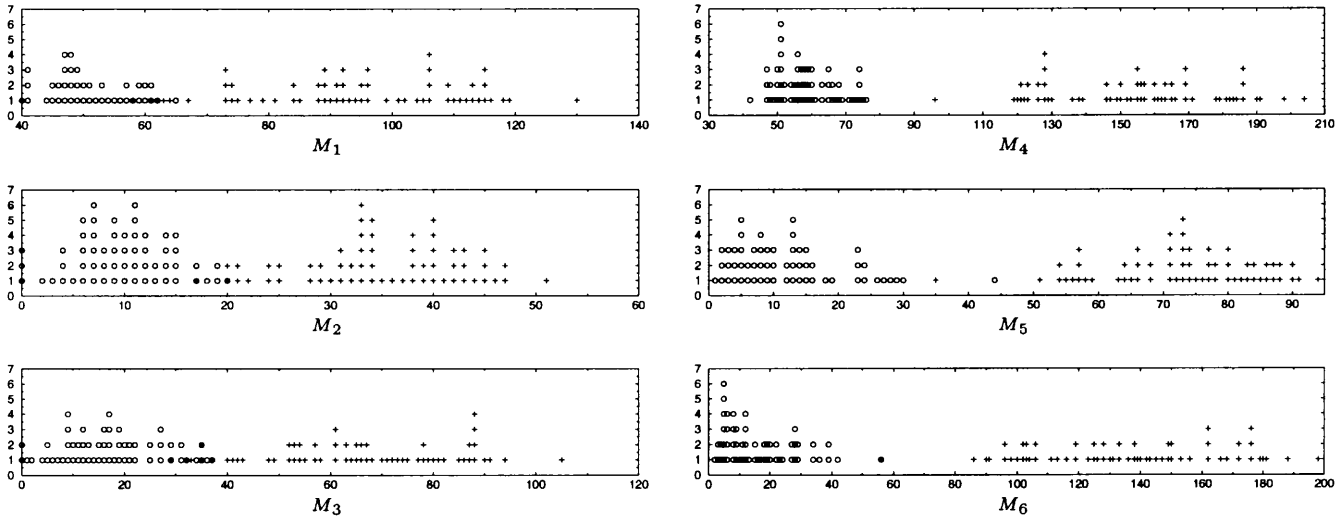

Fig. 8. Histograms of the measures M_k for the image subset S_s.

to be confused by grey level discontinuities caused by hairs. In contrast, M_1, M_2 and M_3 are not that useful for the other two views. The reason for this phenomenon is that our definition of the evaluation area F_1 assumes a view direction more or less straight into the camera. Otherwise, unexpected facial parts with strong grey level discontinuities may fall into this area, and generate false support to the measures. As can be easily seen in Fig. 5, this problem can be caused by the nostrils in the upward view and by the facial border against the background in the side view. Interestingly, the upward and side views have no essential influence to the evaluation area F_2, and the class separability w.r.t. to M_4, M_5 and M_6 therefore remains promising.

Basically, the experimental results above confirmed the expectation that the fundamental assumption behind the measures M_k may be violated in the case of non-frontal views. Their handling needs a more sophisticated clue accumulation to minimise the influences of undesired edge responses.

7. TESTS ON ORL DATABASE

We have tested our glasses detection method on a second image set from the Olivetti Research Laboratory (ORL), which has been used for testing face recognition systems [7,23]. This image set contains 40 people of 10 images each, subject to significant variations in viewing angle. For our tests we selected two frontal images for each person, resulting in a test set of 80 images in total. Among the 40 people, 14 wear glasses and the other 26 do not. Compared to our own test set described in Section 3, the ORL images have

Table 4. Results on the ORL database

	M_1	M_2	M_3	M_4	M_5	M_6
J	8.82	5.32	6.63	10.09	8.29	9.93
D_1	[37,98]	[5,35]	[3,61]	[39,111]	[0,60]	[0,75]
D_2	[86,152]	[26,68]	[44,127]	[113,221]	[45,100]	[71,221]
Overlap (%)	5.00	16.25	11.25	0.00	7.50	2.50

	M_{14}	M_{15}	M_{16}	M_{24}	M_{25}	M_{26}	M_{34}	M_{35}	M_{36}
β	0.7	0.9	0.5	0.7	0.9	0.7	1.4	1.3	0.9
J	12.76	14.09	13.35	12.02	10.57	11.50	12.15	12.21	12.18
D_1	[74,175]	[46,130]	[42,131]	[40,107]	[8,70]	[8,82]	[77,215]	[11,109]	[10,123]
D_2	[180,286]	[144,234]	[139,246]	[110,206]	[75,153]	[84,206]	[213,406]	[110,250]	[124,301]

Fig. 9. Two examples of the ORL database.

a smaller size (92×112 pixels) and a lower contrast. Two examples of this database are shown in Fig. 9. It should also be mentioned that the ORL images only contain the face, and their uniform size has been made possible by clipping the face part from initially acquired images and scaling the resulting face images to the final resolution. In this way, however, some distortions are introduced. For instance, some of the final face images have a smaller (proportional) height compared to the natural aspect ratio of faces. For this reason, the optimisation procedure for determining the dimensional parameters of the evaluation regions defined in Fig. 1 was repeated for this database. The optimal parameters turned out to be $p = 2$, $q = 0.5 - 0.1 = 0.4$, $r = \frac{1}{7} \approx 0.14$ and $s = \frac{1}{3} - 0.12 \approx 0.21$.

The results on the ORL image set are summarised in Table 4. Generally, these results are worse than those on our test set. In particular, the three measures M_1, M_2 and M_3 produce quite large overlap of the two classes. The measures M_4, M_5 and M_6 are better than their respective counterparts, although we are also faced with the problem of overlap here. The overlap can be reduced by using combined measures. In this case, eight measures are free of overlap and only one has a small overlapping area. Overall, the relationship (2) remains true for this image set, indicating the potential of measure combination in improving the performance of glasses detection.

The performance difference between the two image databases seems to be mainly caused by the different lighting conditions. The ORL database has a much weaker diffuse illumination, resulting in lower responses at the glasses in general. Moreover, strong side illumination can be observed in some of the ORL images. In this case, shadow areas arise in the face (e.g. around the nose) that produce false high responses.

8. CONCLUSIONS

As a salient facial feature, the presence of glasses provides a considerable discrimination power for facial image database indexing. In this paper, we have presented an approach towards detecting the presence of glasses in facial images. Based on automatically detected edge locations, measures have been defined for evaluating areas where glasses are expected. In addition, we have investigated combination measures and demonstrated their superior performance to the individual measures. A simulation has been performed to figure out the inaccuracy of eye detection tolerable by the proposed method. It turns out that an eye localisation within the iris' contour is necessary for a sufficient class separation. This accuracy seems to be achievable by the eye detection techniques known from the literature, manifesting the practical applicability of our approach. We have described experimental results on two facial image databases which demonstrated the potential of the approach.

Acknowledgements

T. Perroud helped us implement and test the eye detection algorithm. We want to thank the Olivetti Research Laboratory for providing their ORL facial image database.

References

1. Pentland A, Moghaddam B, Starner T. View-based and modular eigenspaces for face recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition 1994; 84–91

2. Brunelli R, Poggio T. HyperBF networks for gender classification. Proc DARPA Image Understanding Workshop 1992; 311–314
3. Golomb B, Sejnowski T. Sex recognition from faces using neural networks. In: Murray AF (ed) Applications of Neural Networks Kluwer Academic, 1995; 71–92
4. O'Toole AL, et al. Classifying faces by race and sex using autoassociative memory trained for recognition. Proc Annual Meeting of the Cognitive Science Society 1991; 847–851
5. Kwon YH, Lobo N. Age classification from facial images. Computer Vision and Image Understanding 1999; 74(1): 1–21
6. Wiskott L, Fellous JM, Krüger, von der Malsburg C. Face recognition and gender determination. Proc First Int Workshop on Face and Gesture Recognition 1995; 92–97
7. Samaria FS, Harter AC. Parameterisation of a stochastic model for human face identification. Proc 2nd IEEE Workshop on Applications of Computer Vision 1994; 138–142
8. Jia X, Nixon MS. Extending the feature vector for automatic face recognition. IEEE Trans PAMI 1995; 17(12):1167–1176
9. Stringa L. Eyes detection for face recognition. Applied Artificial Intelligence 1993; 7:365–382
10. Benn DE, Nixon MS, Carter JN. Robust eye center extraction using the Hough transform. In: Bigün J, Chollet G, Borgefors G (eds), Audio- and Video-Based Biometric Person Authentication, Springer-Verlag, 1997 1–9
11. Kothari R, Mitchell J. Detection of eye locations in unconstrained visual images. Proc Int Conf on Image Processing, Lausanne 1996; 519–522
12. Nixon M. Eye spacing measurement for facial recognition. SPIE Proc 1985; 575: 279–285
13. Herpers R, Michaelis M, Lichtenauer KH, Sommer G. Edge and keypoint detection in facial regions. Proc Second Int Workshop on Face and Gesture Recognition 1997; 212–217
14. Huang W, Sun Q, Lam CP, Wu JK. A robust approach to face and eyes detection from images with cluttered background. Proc 14th Int Conf on Pattern Recognition 1998; 110–113
15. Chow G, Li X. Towards a system for automatic facial feature detection. Pattern Recognition 1993; 26(12):1739–1755
16. Xie X, Sudhakar R, Zhuang H. On improving eye feature extraction using deformable templates. Pattern Recognition 1994; 27(6):791–799
17. Yuille AL, Hallinan P, Cohen D. Feature extraction from faces using deformable templates. Int Journal on Computer Vision 1992; 8(2):99–111
18. Chellappa R, Wilson CL, Sirohey S. Human and machine recognition of faces: a survey. Proceedings of IEEE 1995; 83(5): 705–740
19. Lee SU, Chung SY, Park RH. A comparative performance study of several global thresholding techniques for segmentation. Computer Vision, Graphics, and Image Processing 1990; 52: 171–190
20. Sahoo PK, Soltani S, Wong AKC, Chen YC. A survey of thresholding techniques. Computer Vision, Graphics, and Image Processing 1988; 41:233–260
21. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans SMC 1979; 9:62–66
22. Achermann B, Bunke H. Combination of face classifiers for person identification. Proc 13th Int Conf on Pattern Recognition, 1996; C:416–420
23. Zhang J, Yan Y, Lades M. Face recognition: eigenface, elastic matching, and neural nets. Proc IEEE 1997; 85(9):1423–1435
24. Schalkoff R. Pattern Recognition: Statistical, Structural and Neural Approaches. John Wiley & Sons, 1992

Xiaoyi Jiang received his BS degree from the University of Peking, China, and a PhD degree from the University of Bern, Switzerland, both in computer science. In 1997 he received the *Venia Docendi* degree (Habilitation) from the University of Bern. Currently, he is a senior lecturer with the Department of Computer Science and Applied Mathematics at the University of Bern. His research interests include computer vision, pattern recognition and computational geometry. He has more than 60 publications in these areas, including a book on three-dimensional range image acquisition and analysis published by Springer-Verlag in 1996.

Michael Binkert is a master student in computer science at University of Bern. He began his studies in 1994 and graduated in October 1999. His interests include image analysis and pattern recognition. In his masters thesis, he is researching the application of shape analysis on unicellular microorganisms called diatoms.

Bernard Achermann is an artificial intelligence and software engineering expert working for the Corporate Information & Technology Unit of Swisscom. Before joining Swisscom in 1998, he received a doctoral degree from the University of Bern for his work in the area of face recognition based on range images. He has published articles on the subject of face recognition methods with hidden Markov models and Hausdorff distance. Currently, he is working in the area of data mining, e-commerce and quality assurance in telecommunications. Furthermore, he has an ongoing interest in object-oriented technologies, Java and C++, and internet applications.

Horst Bunke received his MS and PhD degrees in computer science from the University of Erlangen, Germany in 1974 and 1979, respectively. In 1984 he joined the University of Bern, Switzerland, where he is a full professor in the Computer Science Department. He held visiting positions at the IBM Los Angeles Scientific Center (1989), the University of Szeged, Hungary (1991), the University of South Florida at Tampa (1991, 1996 and 1998), the University of Nevada at Las Vegas (1994), Kagawa University, Takamatsu, Japan (1995), and Curtin University, Perth, Australia (1999). Horst Bunke is a Fellow and one of the current Vice Presidents of the International Association for Pattern Recognition (IAPR). He is associate editor of the *International Journal on Document Analysis and Recognition*, editor-in-charge of the *International Journal of Pattern Recognition and Artificial Intelligence*, and editor-in-chief of the book series on Machine Perception and Artificial Intelligence by World Scientific. He has more than 300 publications, including 20 books and special editions of journals.

Correspondence and offprint requests to: X. Jiang, Department of Computer Science, University of Bern, Neubrückstrasse 10, Ch-3012 Bern, Switzerland. E-mail; jiang@iam.unibe.ch