PREPROCESSING IMAGES OF FACES: CORRELATIONS WITH HUMAN PERCEPTIONS OF DISTINCTIVENESS AND FAMILIARITY

P.J.B. Hancock[1], A.M. Burton[2] and V. Bruce[1]

1. Department of Psychology, University of Stirling, UK
2. Department of Psychology, University of Glasgow, UK

INTRODUCTION

The aim of this work is to further our understanding of how humans process and recognise faces. We are doing this by proceeding in parallel with testing subjects and building computer models. If a model reflects the way that humans process face images, it ought, among other things, to fail in the same way: to find the same faces easy or difficult. One characteristic of human recognition is that of distinctiveness: some faces are never forgotten, others easily lost in a crowd. This paper describes the use of various forms of image processing to see whether they correlate with human perceptions of distinctiveness, memorability and familiarity.

HUMAN MEASUREMENTS

Distinctiveness ratings were obtained for a set of 174 images of young Caucasian males, by asking a total of 34 subjects to rate how easy it would be to recognise them at a station, on a scale of 1-10. Each subject saw half the faces for rating. Following a ten minute distractor task, they were shown the complete set and asked which had been seen earlier. Response was again on a scale of 1-10, where 10 means "certain was seen", and 1 means "certain not seen". The scores for those faces that actually had been presented were averaged across subjects to give a hit score, those for previously unseen faces likewise to give false positive scores. With each group of subjects seeing one half of the face set, scores were obtained for all the faces. The correlations between these scores are shown in table 1, with scatter graphs of the data in Figure 2. While more distinctive faces tend to have higher hit scores, and lower false positive scores, there is an unexpected lack of any negative correlation between hit and false positive scores. The faces that are easy to remember are not necessarily the same as those that are easy to reject as unseen. These findings may be interpreted in terms of two orthogonal factors that have been called memorability and general familiarity (Vokey and Read (8), Bruce et al (3)). Memorability corresponds to the tendency to answer correctly whether a face was present, and is correlated with rated distinctiveness. General familiarity corresponds to a tendency to say yes whether a face was presented before or not, and is, surprisingly, uncorrelated with distinctiveness.

PRINCIPAL COMPONENTS OF FACES

Principal component analysis (PCA) of face images has been suggested as a means of pre-processing for subsequent recognition (Turk and Pentland (7)). Craw and Cameron (4) report an improved recognition performance if faces are morphed to an average shape before PCA. Figure 1 shows the method and results of the shape averaging. The average "shape free" image is much sharper as a result of all the images having their gross features in the same location. This provides a more nearly linear space for the PCA to analyse.

We performed PCA on eye-aligned images of the faces, as used for the ratings experiment, with and without the shape averaging. We also analysed the shape vectors, these being the 35 {x,y} pairs for the control points around each face. Distinctive to a PCA means a large deviation from the mean: if PCA captures the same notion of distinctiveness that humans use, then large component outputs should correlate with the human rating. Figure 4 shows the coefficients of some of the components obtained. These coefficients are commonly known as eigenfaces. The lowest components appear to change rather coarse features, such as face shape and hair length. The later components appear to carry more information about individual identities, consistent with the findings of Abdi et al (1).

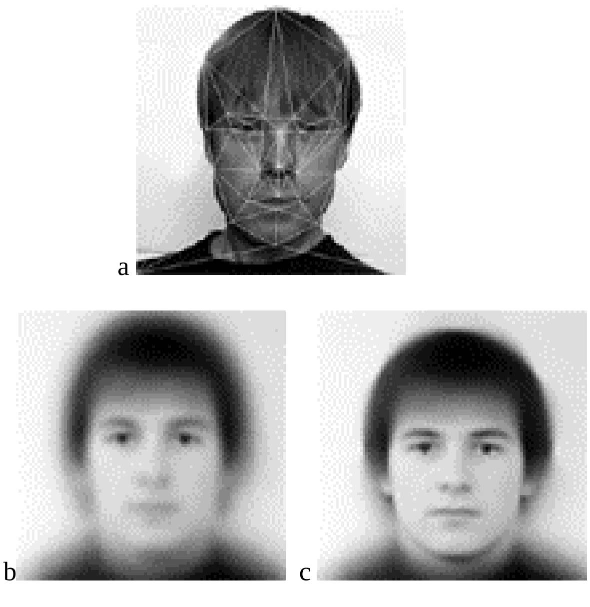


Figure 1. a) Location of the points around each face used for shape averaging, with the morphing triangles

thus defined. b) Simple average of all the eye-aligned face images. c) Average of all the "shape free" images.

	Distinctiveness - Hit	Distinctiveness - False Positive	Hit - False Positive
Set A	0.55	-0.39	-0.17
Set B	0.40	-0.42	0.06
Both	0.49	-0.40	-0.08

Table 1. Correlations between averaged subject responses for the two half groups and combined. Critical values of r for $p<0.05$ are 0.21 for half groups (n=87) and 0.15 for whole set (n=174).

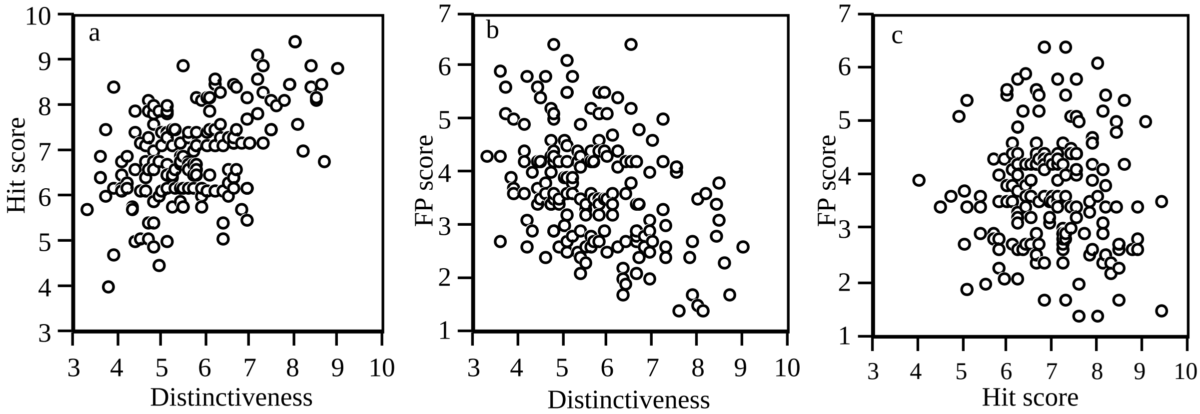


Figure 2. Scatter graphs for values of hit score, false positive score and distinctiveness for the complete set of 174 faces, averaged across 17 subjects each.

Type	Distinctiveness	Hit score	False positive
With shape	0.51	0.42	0.42
Shape averaged	0.48	0.40	0.49
Shape vector	0.48	0.37	0.30
Shape averaged + shape	0.52	0.52	0.48

Table 2. Multiple correlation between principal components and subject ratings, average of 100 random segmentations of the face set, different sets used for generation and test.

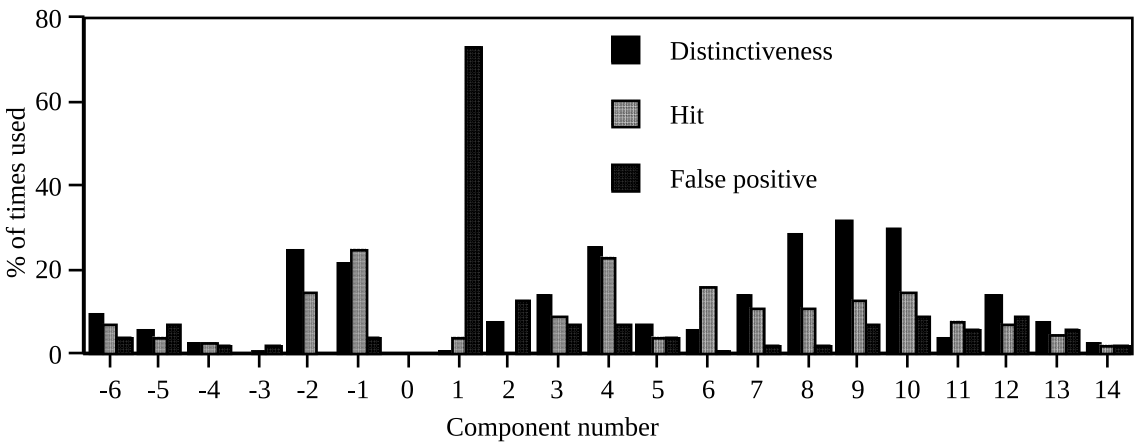


Figure 3. Usage of component outputs in the multiple regression equations, for 100 random segmentations of the face set. Negative numbers refer to the absolute value of the component output, positive numbers to the raw value. Components outside this range were not used significantly.

Components were generated in two ways: performing PCA on the whole set of faces and using the component outputs directly; and splitting the set in half, using one half to generate the PCA coefficients which were then used to analyse the second half of the images. The two methods gave very similar results, a reflection of the size and relative homogeneity of our set of faces. The raw and absolute values of the first 20 component outputs for each face were correlated with the human ratings, both individually and combined in a multiple regression. Table 2 shows results from the multiple regression, average of 100 random segmentations of the set, generating components from one half and testing on the other.

The significance of these results lies in the differences between them. The shape-averaging has little effect on the correlations for distinctiveness and hit score, but significantly increases that with false positive scores. Conversely, the shape vector alone appears to contain no useful information about false positives, the value of 0.3 being the expected multiple-R for this many variables with random data. Performing PCA separately on the shape-averaged images and the shape vector, and taking the first few component outputs from each, gives the best of both. To our knowledge this is the first indication of possible psychological relevance for shape-averaging (Hancock et al (6)).

There are also marked differences in the information carried by individual components. Figure 3 shows the usage of component outputs from the full images in the multiple regression equations. False positive score loads heavily onto the first component output: much of the multiple regression score comes from this one component. Distinctiveness and hit rate show a very different pattern, loading onto the absolute values of the first two components, and the raw values of several other components in the range 4-10. The degree of loading onto raw PC values, rather than their absolute values, was unexpected. Distinctiveness might be expected to lie on both sides of the mean (large or small nose, etc). For the majority of our components, this is not the case.

Figure 4. Top row: the first, second, twentieth and fortieth eigenfaces produced from our set of faces. Bottom row: the same eigenfaces, added to the average face from figure 1b.

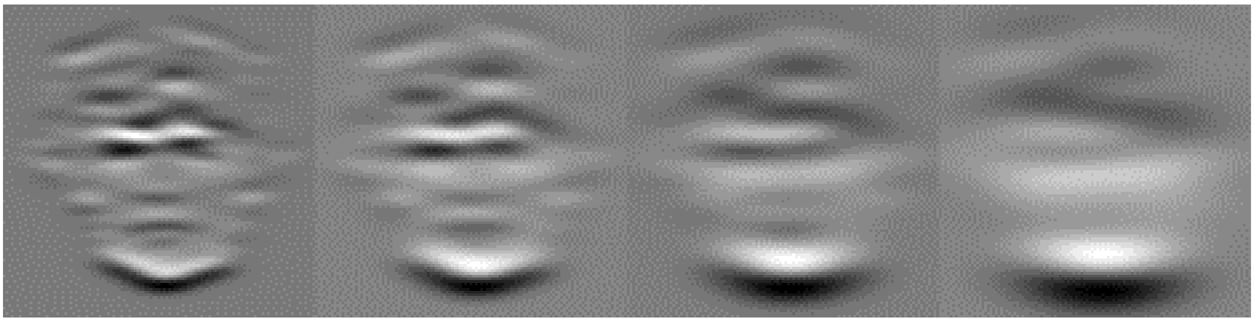


Figure 5. Results of horizontal bandpass filter on one of the faces. Difference of Gaussian filter with standard deviation of 1, 1.5, 2 and 2.5 pixels, acting on 64x64 image.

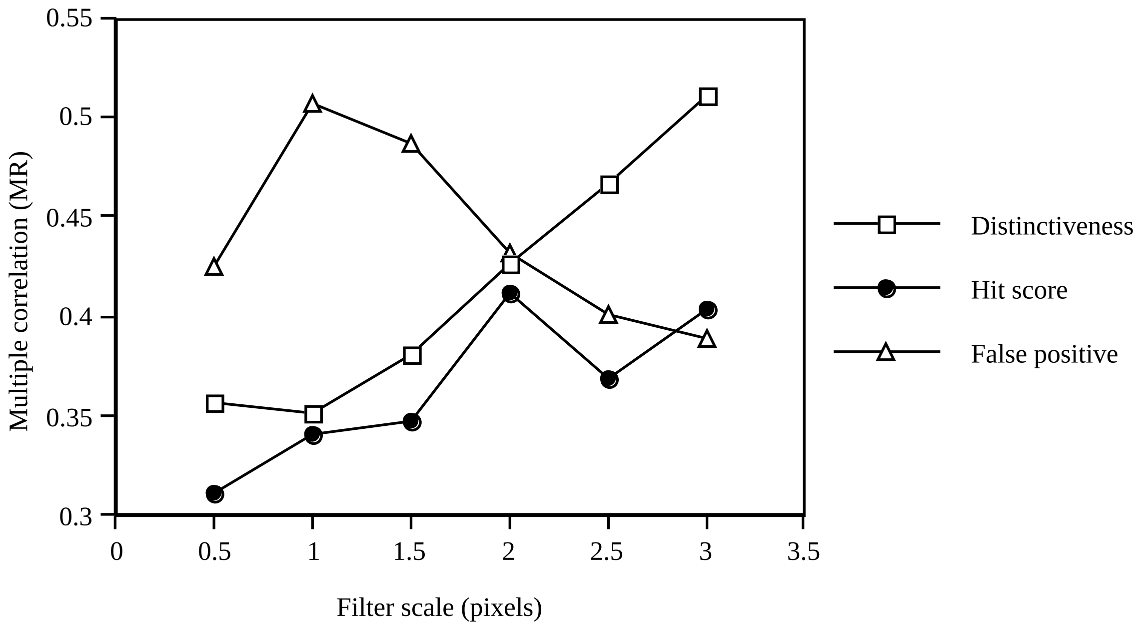


Figure 6. Multiple regression values for the three subject ratings with the raw and absolute values of the first 20 principal components from bandpass filtered face images, average of 100 random segmentations of the face set.

BANDPASS FILTRATION

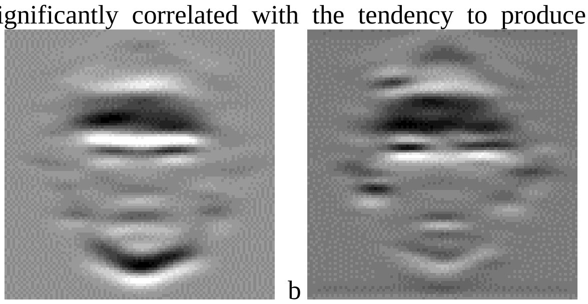
Convolution with horizontally-aligned difference of Gaussian filters was found by Adini et al (2) to minimise variations due to illumination direction, and has also been used by Dakin (5) and Watt (9) to extract experimentally-useful "bar-codes" from faces. We convolved our face images with horizontally-aligned difference of Gaussian filters of varying size, prior to performing PCA. Examples of a face following filtration at a number of scales are shown in Figure 5. Multiple correlations as before with the human data show a very striking pattern of results, shown in Figure 6. At fine scales, there is no correlation with distinctiveness (the recorded value of about 0.3 is at chance for this many variables), but false positives gave a multiple R of 0.51, higher than anything in Table 2. At coarse scales, there is a reversal, with little correlation for false positives, but a multiple R of 0.51 for distinctiveness. The fall-off for false-positives is consistent with the results from the shape vector above: at coarse scales only a general impression of the size and shape of a face remains. Hit score also tends to increase at coarser scales, consistent with the observation that people tend to rely on external features, such as hair, when remembering unknown faces, switching to internal features as the face becomes more familiar.

The trend could not be pursued to coarser scales because the later components became redundant, i.e. the data is less than 20 dimensional when so coarsely filtered. This results in degenerate correlation matrices. Work is in hand to rerun all the tests with fewer components included to allow fair comparison at the extremes of the scale range.

DISCUSSION

The data shown in Figure 2 may have forensic implications. It appears that the notion of some faces being distinctive while others disappear in a crowd is too simplistic. While some faces do seem to be rather anonymous and poorly-remembered, others appear familiar whether previously seen or not. If a suspect has the kind of face that appears familiar, thus falling at the top of Figure 2c where there is a strong tendency to give false positive replies, it might be wise to put less weight on witness identification evidence than if it falls towards the bottom, particularly the bottom right where the ratio of hit rate to false positive rate is highest. At the moment, the only way to ascertain this would be to put the face in amongst a number of others in an experiment similar to the one reported here. Such a procedure would be fraught with difficulties to do with choosing both the other faces and the subject panel and invalidated if pictures of the suspect had been widely publicised. The computer analysis reported here might point towards a way of obtaining such a rating automatically, though it is clear that considerably more work needs to be done before it could have legal standing.

What it is about a face that makes it distinctive or falsely familiar is still not clear. Evidence presented here suggests that the first principal component is significantly correlated with the tendency to produce



a b

Figure 7 a) First eigenface from face images filtered with filter sd 1 pixel. b) Result of filtering the first eigenface from the full images with bandpass filter sd 1.

false positive responses, but it is not obvious why this should be so. This also makes the data from bandpass filtration somewhat puzzling. It appears that coarse scales give information about distinctiveness, which, as noted above, is consistent with the observation that the shape vector of the faces also holds such information. There is little information in the shape about false positive scores; that lies in the shape free image, which carries relatively fine scale "texture" information. This seems consistent with the observed correlation given by fine scale bandpass images. The oddity is that the first principal component, which correlates with false positive scores, appears to yield relatively coarse scale information: see the examples in Figure 4. The explanation may be hinted at in Figure 7. Figure 7a shows the first component obtained from images that were bandpass filtered at a scale of 1 pixel. As with the unfiltered images, it is this first component that is especially correlated with false positive scores. It may be seen that much of the energy lies around the hairline and eyes, and on the chin. Figure 7b shows the result of bandpass filtering the first eigenface from the unfiltered images (as shown in Figure 4) at the same scale. The results are by no means identical, but there is again considerable energy in the region of the hairline and eyes. Why this should predict false positive scores is still a mystery.

CONCLUSION

Our evidence is, therefore, that our subjects are using information at different scales to do different aspects of face processing. Fine scales appear to relate to familiarity, coarser scales to distinctiveness and memorability. We are still exploring the psychological implications of this, but meanwhile hope it may be of use to those more directly interested in engineering applications of face recognition.

ACKNOWLEDGEMENTS
This work was supported by SERC grant number GRH 93828 to Burton, Bruce and Craw.

REFERENCES

1. Abdi H, Valentin D, Edelman B, and O'Toole AJ, 1993, "More about the difference between men and women: evidence from linear neural network and principal component approach" Submitted

2. Adini Y, Moses Y and Ullman S, 1994, "Face recognition: the problem of compensating for changes in illumination direction" Proceedings of ECCV-94, Stockholm, Springer-Verlag.

3. Bruce V, Burton AM and Dench N, 1994, "What's distinctive about a distinctive face?" Quarterly Journal of Experimental Psychology, 47A,119-141

4. Craw I and Cameron P, 1991, "Parameterising images for recognition and reconstruction", Proceedings of the British Machine Vision Conference, 367-370, Mowforth P, editor, Springer Verlag, London

5. Dakin SC, 1994, "The visual representation of texture", PhD Thesis, Department of Psychology, University of Stirling.

6. Hancock PJB, Burton AM and Bruce V, 1995, "Face processing: human perception and principal components analysis", submitted to Memory and Cognition

7. Turk M and Pentland A, 1991, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3, 71-86.

8. Vokey JR and Read JD, 1992, "Familiarity, memorability and the effect of typicality on the recognition of faces", Memory and Cognition 20, 291-302.

9. Watt R. 1994 "A computational examination of image segmentation and the initial stages of human vision" Perception, 23, 383-398.