

WEARABLE FACE RECOGNITION AID

C Iordanoglou, K Jonsson, J Kittler and J Matas

University of Surrey, Guildford, Surrey, GU2 5XH, United Kingdom
email: {J.Kittler,K.Jonsson,G.Matas}ee.surrey.ac.uk

ABSTRACT

The feasibility of realising a low cost wearable face recognition aid based on a robust correlation algorithm is investigated. The aim of the study is to determine the limiting spatial and grey level resolution of the probe and gallery images that would support successful prompting of the identity of input face images. Low spatial and grey level resolution images are obtained from good quality image data algorithmically. The tests carried out on the XM2VTS database demonstrate that robust correlation is very resilient to degradations of spatial and grey level image resolution. Correct prompts have been generated in 98% cases even for severely degraded images.

1. INTRODUCTION

In recent years the computer and sensor technology developments motivated by the market opportunities of mobile communication has resulted in dramatic advances of information technology. The computing power per volume has reached the stage where computationally demanding applications can be realised using wearable electronics. Combined with the concurrent drive to miniaturise camera sensors, information technology is offering the possibility to build smart devices which can act as intelligent aids, helping the user to interact with its environment and other people.

In this paper our interest focuses on a technology which would provide the user with useful information about personal identity. The idea is that a wearable device [8][1] would grab an image of a person the user is approaching and provided the person has been seen by the device before, it would attach a unique name or a short list of name identities to help the user recognising it. The face recognition problem has been studied extensively [3]. In contrast to mainstream work on face recognition we are aiming at a low cost solution which may call for a low-resolution camera, with images acquired by the device being further degraded by additional processing steps of compression and/or transmission over a radio channel.

In this context the paper investigates a face recognition algorithm based on robust correlation reported in [9] [7]. The goal is to evaluate its ability to deliver useful information about a face identity in the case of low quality image data. We shall demonstrate that even with images of severely reduced quality in terms of both spatial and grey level resolution this face recognition approach can provide useful prompting to the user.

The paper is organised as follows. In the next section we describe briefly the face recognition method used. The main body of the paper, Sections 3 and 4 reports on a number of experiments designed to test the robustness of the face recognition method to degradation of face image input. In Section 5 the paper is drawn to conclusion.

2. FACE RECOGNITION BY ROBUST CORRELATION

A face is a surface of a three-dimensional solid having partially deformable parts. Its representation on a two-dimensional field is what a recognition algorithm manipulates to come up with a result. According to [5], "the central factors determining the performance of a face recognition system are the dimensionality and the variance of the set of degrees-of-freedom which are encoded. Ideally, their inter-class variance should be large and their intra-class variance should be small. Therefore different faces should generate face codes that are as different as possible from each other, while different images of the same face should generate very similar codes". Recent investigations of how well this goal is achieved have studied the invariances in face coding schemes under changes in illumination, perspective angle or pose and expression. Their results showed that there is greater variability in a given face over the above transformations, than there is among different faces when these three factors are held constant. This is a fundamental problem in terms of classical pattern recognition addressed in many papers.

The recognition of faces is based on correlating two

images. Previous studies by Brunelli [2] have shown that correlation is not efficient enough for recognition of face images, due to the natural change of appearance of people over time. So for example the use of glasses can result in a failure when recognising a person. It is a challenge to come up with robust methods, which are able to perform well with changes in fundamental figures. Robust correlation identifies the best matching pair of images while ignoring any gross local dissimilarities which may exist in a pair of images of the same person due to for instance glasses, beard, etc. Robust statistics, in general, provides the ability to approach structures by intelligently weighting data. In essence, robustness is achieved by setting the contributions of outliers to a correlation measure to a constant. The data dependent weighting of contributions is controlled by a function called robust kernel. A cut-off distance essentially defines the pixels, which are considered important during recognition. The smaller the intensity difference between corresponding pixels, the larger the kernel response. For differences greater than the cut-off distance, the correlation response is set to zero. Therefore, the robust correlation kernel is positive, taking values between 0 and 1 for normalised images.

The kernel adopted in this work is defined by a truncated quadratic function [6]. The cut-off distance can take values between 0 and 255. Varying the cut off, different kernels are obtained and thus different performances. Practically speaking, a value close to 0 makes the kernel more sensitive to outliers, while a value close to 100 simulates the performance of correlation. In the experimental section of the project, various cut-off distances were tested to find an optimal value.

3. EXPERIMENTAL METHOD

Images from the multimedia database XM2VTS[10] were used for the experiments. This is a multimodal database of 295 people. For each person there are four sessions and two shots per session. We have used the first 200 people from the database in our experimentation. For each person two frontal face images were extracted from the sequences recorded during each session. We used images from session 1 & 2 for training. The testing set included probe vectors corresponding to 100 faces, two for each person taken from sessions 3 & 4. Preregistration of target and gallery images was performed manually within a loop. At each step of the loop a test image was read, degraded and photometrically normalised. This was the image to be correlated with all images of the gallery within a second loop. Within this loop gallery images were photometrically normalised, as well, before the kernel application. For

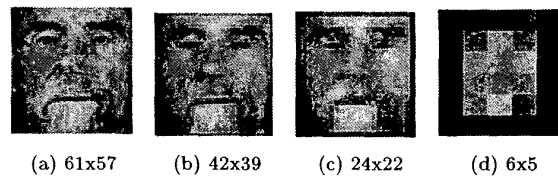


Figure 1: An example of an image (a), subsampled at different scales. The resulting images (a), (b) and (c) exhibit the apparent effect of pixelisation

every probe, 200 robust correlations were computed (the size of model database) and the ones with the highest response were ordered to form a response vector. In order to keep account of the recognition statistics, cumulative match scores were calculated for each probe image. Cumulative match scores show for what fraction of the probes the correct answer is among the top k ones (k denotes the rank of the image to be identified). Ideal recognition rank is 1.

3.1. Image degradation

Spatial Quantisation Noise: Each pixel in an image corresponds to a small area of that image. An ideal representation of an object can be achieved by choosing the image pixel area to be small enough, i.e. by employing a high-density pixel matrix. The process of breaking the continuous image into a grid of pixels is called spatial quantisation. The spatial resolution affects our perception of the quality of the image. A picture consisting of 16384 pixels arranged in a 128 x 128 grid is more easily recognised than its degraded version with 169 pixels arranged in a 13 x 13 grid. When low-resolution cameras are employed, the captured images exhibit an apparent effect of pixelisation. The checkerboard effect, as it is known, is a form of degradation - loss of information. This effect can practically be simulated by sub-sampling a high resolution image.

In this work we have subsampled the original face images of 61 x 57 resolution in three steps to the lowest resolution of 6 x 5. The results of the subsampling are shown in Figure 1.

Grey level quantisation noise: The sensitivity of the camera determines how efficiently light reflected from the imaged scene is recorded thus defining the quality of the captured image. In the case where a low-sensitivity camera is used, the resulting image is subject to a form of degradation. The amount of retained information is also dependent upon the process of quantising the brightness levels of each pixel into a grey level value. Quantised grey levels are represented

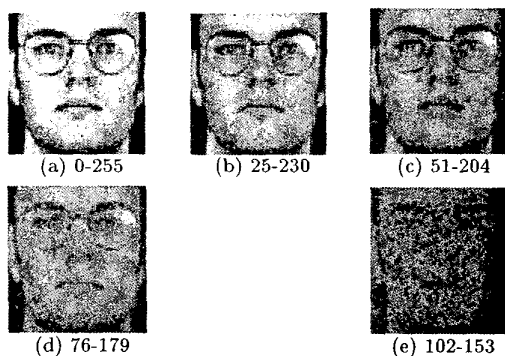


Figure 2: Original image (a), and images with (b) 205, (c) 153, (d) 103 and (e) 51 greylevels

by binary digits, with the number of grey levels being a power of 2. Reducing the bits per pixel different grey scale resolutions of the image are obtained. 5bits correspond to 32 grey levels, 3bits to 8 grey levels and 1bit to 2 grey levels. The effect of these reduced quantisations is illustrated in Figure 2. Note that in the extreme cases the reduction in the image grey levels produces false contouring. As far as face images are concerned, the resulting lack of information produces a less detailed image (or noisy image). One would expect that this would make the recognition more difficult.

4. EXPERIMENTAL RESULTS

4.1. Recognition rate as function of spatial quantisation

In this experiment the original 61 x 57 images were subsampled as described in the previous section. In order to minimise aliasing effects, a 3 x 3 low pass filter was applied to the original image before interpolation. A practical problem was that the representation matrix of the probe is of reduced size, and a direct correlation with the model image matrix in the database was impossible. The easiest way to deal with it was to subsample the database images as well with the same factor as the probe. The results of the recognition procedure on the degraded images are shown in Figure 3. The results are parametrised by the cut off threshold which for value 60 and above would correspond to standard correlation.

The graphs suggest similar recognition rates for images of resolution (a), (b) and (c), scoring 90%. For images of resolution (d), the recognition rate falls to above 70% and rises up to 95% within rank 10. In this particular experiment only a minimal dependence

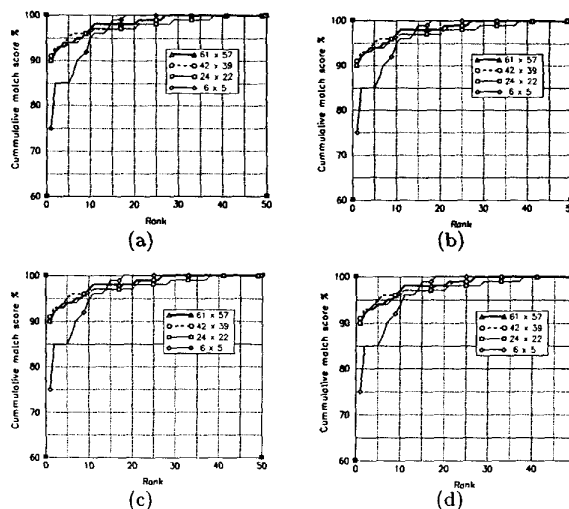


Figure 3: Recognition varies smoothly over (a), (b), (c), (d) for cut-off values 30, 40, 50 and 60

of the performance on the cut off threshold was observed. Surprisingly, the lowest resolution probes reach the recognition score of 100% (within rank 20) before the others. This interesting behaviour is due to the way the recognition was performed. Subsampling both probes and models results in images of low quality but they still appear to encode the necessary discriminatory details of faces. Clearly if a low quality probe was supersampled and tested against the original gallery the conclusions might be different. However, with the proposed symmetric degradation of the gallery images the performance of the algorithm appears not to suffer.

4.2. Recognition rate as function of grey level reduction

The admissible range of pixel intensity values is between 0 and 255. Different degradations can be produced by restricting the picture grey level values to a smaller range, thus reducing the number of grey levels characterising each image. In this experiment the grey levels of the original image were mapped to intensity values corresponding to the central grey levels of the intensity range as shown in Figure 4. Note that from the practical point of view, only the ranges which are powers of 2 would have impact on the wearable aid complexity.

With this form of degradation the recognition rate stays at the same level no matter how severe the grey level reduction is. This shows once again a surprising robustness of the face recognition method.

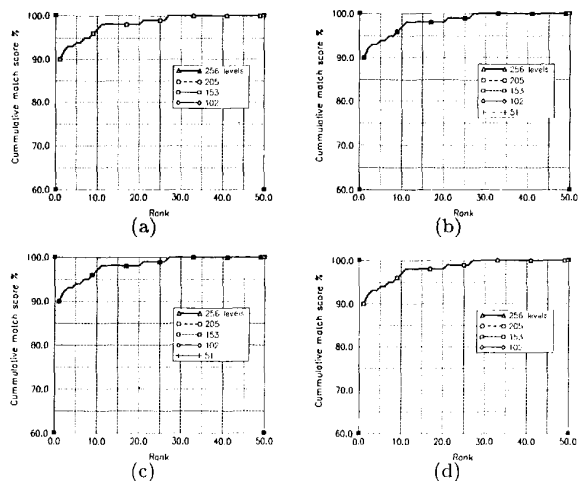


Figure 4: Greylevel reduction does not seem to affect recognition at any of the mapping levels produced in all cases of (a) 30, (b) 40, (c) 50 and (d) 60 cut-offs.

5. DISCUSSION AND CONCLUSIONS

The results obtained demonstrate a remarkable resilience of the face recognition algorithm to degradations in spatial and grey level resolution. The recognition performance reaches 91% for rank 1 recognition, rising up to around 98% for rank 10 which would be quite a favourable rate even for high-performance recognition systems. It is surprising that the recognition rate of degraded probes is higher than recognition of the non-degraded ones, in the rank range 0 to 25. This is an interesting behaviour, which elucidates the robustness of the algorithm with respect to degradations. This finding is consistent with results reported in psychophysical literature [4]. Concerning the cut-off distance, a value of 50 makes the kernel more robust with a performance of 91% for rank 1 to 98% for rank 10. A value of 40 results in responses ranging from 92% to 96% respectively. The former provides an overall higher performance, while the latter offers an individual maximal performance. The use of one or the other is relative to which rank is preferred.

The wearable face recognition system based on the investigated algorithm has the potential to provide valid information to the user. The higher the recognition rate, the more useful the results will be. This is to say that instead of providing minimal information of certain validity, the system could provide more ambiguous information with a higher probability of being truthful. Usually humans remember faces but not names. If the user of the system is provided with a list of names, he

will be able to identify the person he is interested in. A list of 10 names corresponding to the 10 best matches, would give the essential hints to the user. At the same time a higher performance is guaranteed. From an operating point of view, an optional setting could be realized, allowing the user to choose a prompt list size and thus the performance of the system.

Another point to note is that the robust correlation exhibits, on average, a factor of 3% higher performance compared to normal correlation. This factor is essential for systems that provide highly reliable information. A recognition rate of 90% instead of 88% justifies the choice of robust correlation over the standard method.

In summary the experimental study demonstrated the feasibility of a wearable face recognition aid based on the robust correlation algorithm.

REFERENCES

- [1] W Barfield and K Baird. Issues in the design and use of wearable computers. *Virtual Reality*, 3:157-166, 1998.
- [2] R. Brunelli and S. Messelodi. Robust estimation of correlation with applications to computer vision. *Pattern Recognition*, 28:833-841, 1995.
- [3] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):704-740, May 1995.
- [4] N Costen, D Parker, and I Craw. Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics*, 58:602-612, 1996.
- [5] J. Daugman. Face and gesture recognition: Overview. *PAMI*, 19(7):675-676, July 1997.
- [6] F. R. Hampel, E. M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics*. John Wiley, 1986.
- [7] K Jonsson, J Matas, J Kittler, and S Haberl. Saliency-based robust correlation for real-time face registration and verification. *Proc British Machine Vision Conference*, pages 44-53.
- [8] S Mann. Humanistic computing: "wearcomp" as a new framework and application for intelligent signal processing. *PROCEEDINGS OF THE IEEE*, 86:2123-2151, Nov 1998.
- [9] J. Matas, K. Jonsson, and J. Kittler. Fast face localization and verification. *Image and Vision Computing*, 17(8):578-581, June 1999.
- [10] K Messer, J Matas, J Kittler, J Luetttin, and G Maitre. XM2VTSDB: The extended M2VTS database. In R. Chellappa, editor, *Second International Conference on Audio and Video-based Biometric Person Authentication*, pages 72-77, Washington, USA, March 1999. University of Maryland.