Improving Face Localisation using Claimed Identity for Face Verification

Scott Lowther, Chris McCool, Vinod Chandran and Sridha Sridharan

Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane 4001, Australia.

s.lowther@qut.edu.au, c.mccool@student.qut.edu.au, v.chandran@qut.edu.au & s.sridharan@qut.edu.au

Abstract **- Face localisation is integral to any face verification system and if a face is not localised correctly then verification performance will suffer. This paper presents a technique which localises faces using a holistic approach incorporating the use of the claimed identity. Principal Component Analysis (PCA) is applied over windows across multiple scales of an input image and a face is localised by fusing reconstruction error with a soft intra-class verification decision. Results on the BANCA dataset [1] show that the technique is accurate and robust to large variations in environmental conditions. A variant of this technique was used in our system for the ICPR 2004 face verification contest [2].**

Index Terms – Face Localisation, Face Detection, Principal Component Analysis.

1. Introduction

Face localisation is integral to any face verification system. Inadequate localisation and segmentation in a face verification system will undoubtedly lead to poor verification performance regardless of the discriminative power of the verification algorithm. Numerous techniques have been proposed in recent years to address the problem of detecting and localising face images from input images. A comprehensive survey of recent face detection and localisation techniques is presented in [3]. Current techniques use generic methods to find all candidate faces in an input image.

An area currently overlooked in face detection and localisation research is the use of the subject's claimed ID. In a typical face verification application, the subject's claimed ID is known before localisation is attempted. The prior knowledge of the subject's claimed ID can be used to improve face localisation accuracy. Note that in verification systems accurate localisation is only important for true client accesses; inaccurate localisation for imposter access will lead to true rejection.

The algorithm for face localisation presented in this paper is based predominantly on Principal Component Analysis (PCA) using both Distance From Feature Space (DFFS) and Distance In Feature Space (DIFS). These algorithms are presented in Section 2. A discussion of the experimental procedure and the corresponding results are presented in Section 3. Conclusions and a brief discussion of possible future work are presented in Section 4.

2. Algorithm

2.1. PCA for Localisation

PCA, first introduced into the field of face recognition in [4], performs dimensionality reduction by applying eigen-decomposition to the covariance matrices of a set of training images from which the *M* most descriptive components are chosen. The initial vector space, obtained through eigen-decomposition, is described by *N* (where the input image contains *N* pixel values) eigen-values and eigen-vectors, and as such can represent an image (1). The eigen-values obtained represent the squared variance of each eigen-vector, therefore when it comes to choosing the most descriptive eigen-vectors the *M* eigen-vectors with the greatest eigen-values are chosen. The dimensionality reduction leads to an estimate of the input image (2). This introduces some error (3) and leads to the classic trade off between dimensionality reduction and the error in representation.

$$A = \sum_{j=1}^{N} \alpha_j v_j + \overline{V} \tag{1}$$

$$\overline{A} = \sum_{j=1}^{M} \alpha_j v_j + \overline{V} \tag{2}$$

$$\varepsilon = J - \overline{A} \tag{3}$$

where J is the concatenated input image vector, A and $\bar{A}$ are the reconstructed input image vectors for all eigen-vectors and the M greatest eigen-vectors respectively, α are the coefficient weightings, υ are the eigen-vectors, $\overline{V}$ is the average face and ε is the error in representing J with $\bar{A}$.

Two methods are used to analyse the vector space produced by PCA; DFFS and DIFS. The relationship between these two metrics is illustrated in Figure 1. DFFS represents the error introduced when an image is projected into the vector space (3), and DIFS describes where in the vector space the projected image lies with respect to other projected images of the same class. It was shown in [5] that DFFS and DIFS can be used to conduct object localisation.

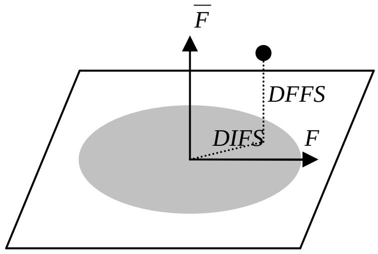


Figure 1: Illustration of DFFS (in the $\overline{F}$ space) versus DIFS (in the F space) [5].

To conduct object localisation using DFFS the input image is projected into the vector space and reconstructed to form Ā. The difference between J and $\bar{A}$, the DFFS or error in reconstruction, is then calculated and the region which produces the least error is then chosen as the objects location. Object localisation using DIFS exploits the fact that objects of the same class lie close to one another in the vector space. In order to calculate DIFS the feature vectors from all training instances of a class are modelled and the derived model is used to obtain a distance measure. There are several ways to model the feature vectors including; average Euclidean distance, multi-variate Gaussian Mixture Models (GMMs) and neural networks (NNs).

2.2. Image Normalisation

Before searching for the location of a face, the input image is normalised to improve localisation in the presence of adverse lighting conditions. Local mean normalisation is applied across the entire image. The mean of a fixed window is subtracted from the centre pixel and the result is stored as the new pixel value in the normalized image (4). Pixels lying outside the image are ignored in the mean calculation for border regions as shown in (5) and (6).

$$N_{i,j} = I_{i,j} - \frac{\sum_{k=i-w}^{i+w} \sum_{l=j-w}^{j+w} I_{k,l}}{P} \tag{4}$$

$$k = \begin{Bmatrix} 1, & i-w<1 \\ i-w, & 1 \le i-w \le m \\ m, & m < i-w \end{Bmatrix} \tag{5}$$

$$l = \begin{Bmatrix} 1, & j-w<1 \\ j-w, & 1 \le j-w \le n \\ n, & n < j-w \end{Bmatrix} \tag{6}$$

Where I is the input image, i, j are image pixel indices, k, l are local region indices, w is half the local mean window size, m and n are the width and height of the input image respectively and P is the number of pixels used in calculating the mean (usually $\{2w+1\}^2$*, but smaller when k and j exceed the input image dimensions).*

This normalization approach requires no prior knowledge of facial structure and minimal computational overhead as it can be applied globally to each image scale before searching begins. This normalization approach increases localisation performance as detailed in the results. Figure 2 depicts an example image before and after image normalisation.

(a) (b)

Figure 2: Example image before and after image normalization. (a) Greyscale image of an in house database subject. (b) The corresponding image after image normalization.

2.3. Intrapersonal DIFS modelling

The large variation in face images and limited training data in current face image databases often leads to weak DIFS models. Intrapersonal variation [6] can aid in DIFS modelling by reducing the problem of modelling all face vectors in the face feature space to instead modelling only the intra-class variation of face vectors within the face feature space. To create an intrapersonal DIFS model, the PCA space for both DFSS and DIFS calculations is first created from face training images normalised as described in section 2.2. All permutations of each individuals training feature vectors in the PCA space are then subtracted and used to train an intrapersonal GMM space. During localisation, the feature vector from each window location is subtracted from the feature vector of an enrolment image of the claimants claimed ID and classified with the intrapersonal GMM. If the window feature vector corresponds to a face image of the same individual as the enrolment image, the intrapersonal model should provide a strong response. Such an approach has the added benefit of utilising claimed identity in the localisation process.

One difficulty with the use of DFFS and DIFS is accurately fusing the two metrics for improved localisation performance. Each is calculated in a different manner, has different scales, ranges and signs. Weighted summation of the range normalised DIFS and DFFS values via thresholds determined from training were used in initial experiments and gave good performance.

2.4. Coarse Search

A coarse search is performed to find an approximate scale and position of a face in a normalised image and is calculated for a number of specified scales. For each scale, the image is normalised by the procedure described in Section 2.2 and the resulting image is searched with a sliding window for the presence of a face. Low activity areas are first removed by calculating the horizontal and vertical standard deviation for each window location and applying a threshold to determine if enough variation exists for further processing. Both the DFFS and DIFS are then fused in calculation of the best position per scale for a face. Figure 3 depicts a block diagram of the operation of the proposed algorithm for coarse face localisation.

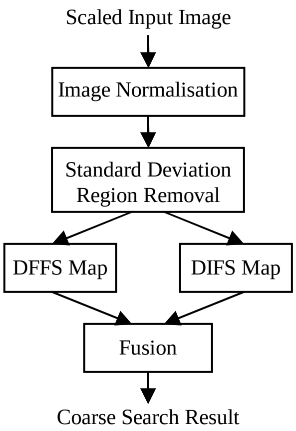


Figure 3: Block diagram of the proposed coarse face localisation algorithm.

The PCA space for DFFS and DIFS calculations is derived from normalised training face images and the highest ranked dimensions retained. Due to the normalisation procedure, border pixels in the training images will be calculated differently to face border pixels in the input image since the input image is normalised globally for each scale instead of for each window location. However, normalising for the entire image is less computationally intensive and should have minimal impact on localisation accuracy. Each window image is converted to a concatenated vector and the feature vector determined according to (7). Once extracted, the DFFS can be calculated as the sum of squared error between the original window vector and the reconstructed image vector as shown in (8).

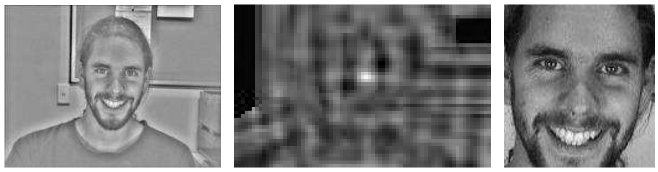
$$\gamma = V \bullet (J - \overline{V})^T \tag{7}$$

where γ is the resulting feature vector, V is the PCA space, $\overline{V}$ is the average face and J is the concatenated window vector.

$$DFFS = \sum (J - (\gamma^T \bullet V + \overline{V}))^2 \tag{8}$$

To reduce false detection and improve localisation accuracy, DIFS is also calculated for each window. DIFS is calculated by subtracting the feature vector γ from the client's enrolment feature vector and projecting this difference vector into the intrapersonal GMM space. The intrapersonal GMM space should be capable of determining if the difference vector produced from subtracting two feature vectors is likely to correspond to the difference between two vectors of the same individual.

Each metric is scaled in a range dictated from thresholds extracted during training and the resulting DFFS map subtracted from the DIFS map. This is repeated for each scale and the best result overall is chosen. Figure 4 depicts an example of coarse image searching.



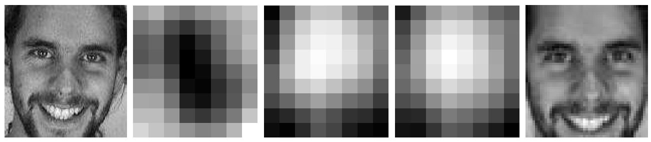
(a) (b) (c)

Figure 4: Example results of coarse searching. (a) Original greyscale image after image normalization. (b) Final DIFS and DFFS maps after low standard deviation removal. (c) Coarse face localisation result.

2.5. Fine Search

The approximate face location and scale obtained from the coarse search is then searched using finer scales and in-plane rotations. As with the coarse search, image normalization is applied to each scale to aid in the detection process. For each scale, the image is first resized then a number of rotations are applied. Selection of the best position within each scale and rotation is performed by fusing the range normalized DFFS and DIFS maps and selecting the maxima. Selection of the best scale and rotation is then performed by which produces the least DFFS error.

The fine search can be repeated again to further improve results. Figure 5 depicts an example of fine image searching.



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

Figure 5: Example of fine search operation (a) Original coarse face image. (b) DFFS map. (c) DIFS map. (d) Fused map. (e) Final localisation result.

3. Experiments

Experimentation was performed on the recently released BANCA dataset [1] using protocol P with the models trained on group 2 and tested on the group 1 evaluation set (2990 test images). The BANCA dataset contains large variation in imaging conditions and provides a good platform for benchmarking of localisation performance for verification tasks. The PCA space was developed using all face images from group 2 at a resolution of 55 pixels high by 51 pixels wide and the first 50 principal components retained. The DIFS model was generated from all client test images from group 2 and modelled as a 4 mixture, 50 dimensional GMM. For comparing features for DIFS, only the first training image of each client in group 1 is used. Only 3 search scales were used for coarse searching while fine searching consisted of a maximum of 7 scales and a maximum of 10 rotation angles.

The first experiment was designed to evaluate the performance of the technique for both true client access images and imposter access images. For comparison, results for our previous technique described in [7] are also provided as well as results obtained without the use of simple image normalisation. Estimated eye locations were determined simply by calculating the position of eye centres in the training data with respect to the centre training images and applying the rotations and scales back to the original input image. Results are shown in table 1. Eye localisation error was calculated according to (9) [8] and determined to be correctly localised if the localisation error was less than 0.25.

$$e_{eye} = \frac{\max(d_l, d_r)}{d_{eye}} \tag{9}$$

where e_{eye} is the eye localisation error, d_l is the distance between true and estimated left eye position, d_r is the distance between true and estimated right eye position and d_{eye} is the distance between the left and right true eye positions.

Technique	**Client images**	**Imposter images**
New technique without image normalisation	85.7%	57.9%
New technique with image normalisation	96.2%	90.1%
Previous technique [7]	56%	55.3%

Table 1: Results of eye localisation experiments on the BANCA dataset for both true client and imposter accesses.

Since localisation errors can often occur from out-of-plane rotations, poor image quality or imposter accesses, all of which create difficulty for verification techniques, it is more meaningful to show localisation performance as the result of verification accuracy. Showing localisation performance via verification performance is also advantageous since pupil locations themselves can vary relative to the face image. Table 2 shows the comparative verification performance difference between automatically extracted face images using the results of the localisation method described in this paper and manually extracted face images using hand marked eye locations. A simple PCA Euclidean distance verifier was employed for verification performance calculations.

Localisation Technique	**Verification Equal Error Rate**
Automatic	34%
Manual	27.8%

Table 2: Results of PCA verification on both automatically and manually extracted face images.

4. Conclusions

This paper has described a simple but accurate method for face and eye localisation utilizing claimed identity for verification tasks. Results, particularly when used in conjunction with a verifier, indicate strong performance for verification systems. While localisation accuracy for imposter accesses is considerably lower than for true client accesses, this should have minimal effect for verification tasks as inaccurate localisation in the case of imposter access often leads to true rejection. The accuracy of the proposed technique is significantly improved by use of a very simple, yet effective image normalisation technique requiring no prior knowledge of lighting conditions or facial structure. Future research will investigate the use of a simple skin classifier to aid in both improving localisation performance and reducing computational requirements.

Acknowledgements

This research was supported by the Office of Naval Research (ONR), USA, under Grant Award No: N000140310663 and by the Australian Research Council (ARC) through Discovery Grants Scheme, Grant DP0452676, 2004-6.

References

[1] BANCA; *http://www.ee.surrey.ac.uk/banca*

[2] BANCA; *http://www.ee.surrey.ac.uk/banca/icpr2004/*

[3] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, 2002.

[4] M. Turk and A. P. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-96, 1991.

[5] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 696-710, 1997.

[6] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian Similarity Measure for Direct Image Matching", proc. of the Thirteenth International Conference on Pattern Recognition, pp. 350-358, 1996.

[7] D. Butler, C. McCool, M. McKay, S. Lowther, V. Chandran, and S. Sridharan, "Robust Face Localisation Using Motion, Colour & Fusion", proc. of the Seventh International Conference on Digital Image Computing: Techniques and Applications, pp. 899-908, 2003.

[8] O. Jesorsky, K. Kirchberg, and R. Frischholz, "Robust Face Detection using the Hausdorff Distance", proc. of the Third International Conference on Audio and Video Based Biometric Person Authentication, pp. 90-95, 2001.