Frontal View Human Face Detection and Recognition

L. S. Balasuriya and Dr. N. D. Kodikara
Department of Statistics and Computer Science, University of Colombo, Sri Lanka

ABSTRACT

This paper is about an attempt to unravel the classical problem of automated human face recognition. A near real-time, fully automated computer vision system was developed to detect and recognise expressionless, frontal-view human faces in static images. In the implemented system, automated face detection was achieved using a deformable template algorithm based on image invariants. The natural symmetry of human faces was utilised to improve the efficiency of the face detection model. The deformable template was run down the line of symmetry of the face in search of the exact face location. Once the location of the face in an image was known, this pixel region was extracted and the test subject was recognized using principal component analysis, also known as the eigenface approach.

1.0 INTRODUCTION

While research into face recognition dates back to the 1960's, it is only very recently that acceptable results have been obtained. Face recognition is not only one of the most challenging computer vision problems but also has many commercial and law enforcement applications. Mugshot matching, user verification and user access control, crowd surveillance and enhanced human computer interaction all become possible if an effective face recognition system could be implemented.

The problem of automated face recognition is generally addressed by functionally dividing it into face detection and face recognition. Before actual face *recognition* is possible, one must be able to reliably find a face and its landmarks in an image. This process, which is called face detection, is essentially a segmentation problem and in practical systems most of the effort goes into this task. In fact, recognition based on features extracted from these facial landmarks is only a minor last step. Most implemented face detection systems use an example based learning approach to determine whether a face is present in a particular pixel 'window' [1]. A neural network or some other classifier is trained using supervised learning with 'face' and 'non-face' examples, thereby enabling it to classify a particular pixel region in an image as a 'face' or 'non-face'. Unfortunately, while it is relatively easy to find face examples, how would one find a representative sample of images which represent non-faces? Therefore face detection systems using example based learning need literally thousands of 'face' and 'non-face' example images for effective training[2]. In this study we used a deformable template to detect the image invariants of a human face. This technique did not need the extensive training of a neural network based approach yet yielded a perfect detection rate for frontal-view face images with a reasonably plain background.

Most of the pioneering work in face recognition was done based on the geometric features of a human face[3], although Craw et. al.[4] did relatively recent work in this area. This technique involves computation of a set of geometrical features such as nose width and length, mouth position and chin shape, etc. from the picture of the unknown face we want to recognise. This set of features is then compared with the features of known individuals and the closest match is found. The main disadvantage of this recognition model is that the automated extraction of these geometrical features is very hard and is therefore more suitable for a system where facial features are extracted manually [5],[6]. This is not the ideal model for a fully automated face recognition system. Face recognition based on geometrical features is also very sensitive to the scaling and rotation of a face in the image plane[7] and therefore would not be as robust as other recognition models.

In face recognition, we attempt to find the closest known face to the unknown face presented to the system. A template matching strategy was used for face recognition in this study. Here, whole facial regions or pixel areas are extracted and compared with the stored images of known individuals and the closest match is found. While the simple technique of comparing grey-scale intensity values for face recognition has been used in the past [8], there are far more sophisticated methods of template matching for face recognition which involve extensive pre-processing and transformation of the extracted grey-level intensity values. The principal component analysis or eigenfaces approach used in this study is such a strategy.

2.0 FACE DETECTION

While there is a great deal of variation among grey-scale human face images, there are several invariant grey-scale regions present. For example, the eye-eyebrow area seems to always contain dark intensity gray-levels, while nose, forehead and cheek areas contain bright intensity grey-levels. The implemented face detection system is able to identify these characteristics and thereby detect a frontal view human face.

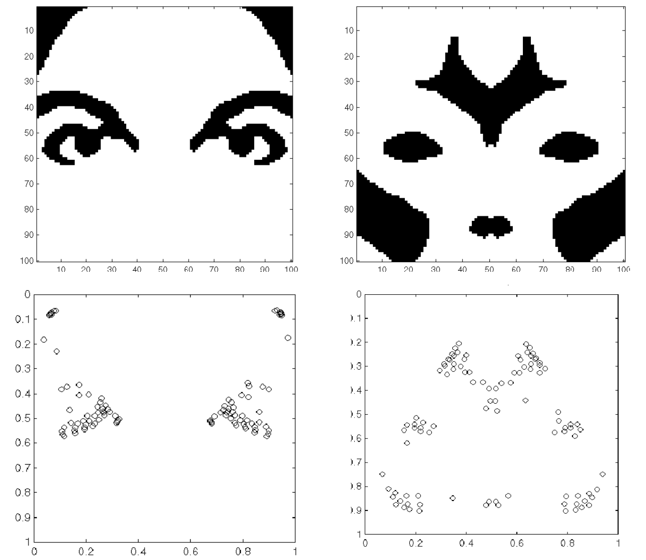

Figure 1. Basis for dark and bright intensity invariant templates (above), and the actual templates that were used in the implemented face detection system(below). We discovered that attempting to detect a facial area that was slightly above the norm yielded more accurate detections and face segmentations. This is probably because of the clear divisions of the bright intensity invariants by the dark intensity invariant regions in this facial area.

These dark and bright greyscale intensity invariant regions were subjectively identified and fed separately into a Kohonen Feature Map with an input space neighbourhood and node sensitivity, thereby creating two network weight topologies that could be used as A-units for a perceptron. The deformable template was implemented by turning the weights of the perceptron's A-units into array indexes, which enabled the system to efficiently extract the gray level intensities from the required positions of the potential face segment. A heuristic was then calculated on the 'faceness' of the segment. Finally, the system chose the pixel area with the highest heuristic as the best possible face segment in the image.

Since there are potentially almost an infinite number of possible locations of a face in an image, an exhaustive search for a face would be computationally demanding. Therefore, the natural symmetry of faces was utilised to improve the efficiency of the face detection model. The correlation of the pixel regions on either side of the potential line of symmetry was calculated and the location with the best vertical symmetry was determined. The deformable template was then run down this line of symmetry in search of the exact face location.

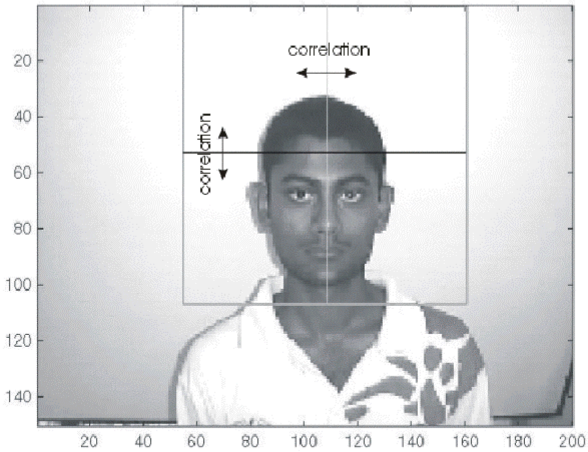

Figure 2. Pixel areas are sampled from left to right on the upper part of a test subject's face image in search of the line of symmetry. This will be in an area with high vertical symmetry yet low horizontal symmetry. The heuristic that was used was the vertical correlation coefficient minus the horizontal correlation coefficient. The area with the highest heuristic value was determined to contain the line of symmetry.

Figure 3. The deformable template travels vertically downwards several times along the test subject's line of symmetry, gradually reducing in size and calculating the 'faceness' heuristic of the sampled pixel area. The pixel area with the highest 'faceness' value (lower right) was judged to contain the best segmentation of subject's face.

Occasionally the best heuristic value did not coincide with the best face location. Therefore the system was designed to examine several of the high 'faceness' pixel areas for correlation with the average human face image (average face of the test subjects in this study). Calculating correlation is computationally expensive and therefore cannot be used as the sole face detection technique. However testing for correlation was useful when paired with a deformable template which reduces the search space (for a face) from almost an infinite number of locations in an image to a few possibilities. This two-tier detection approach enabled the system to be fast as well as accurate.

'Faceness' heuristic	Location		
	x	y	width
978	74	31	60
1872	74	33	60
1994	75	32	58
2418	76	34	56
2389	79	32	50
2388	80	33	48
2622	81	33	46
2732	82	32	44
2936	84	33	40
2822 Actual Face location	**85**	**58**	**38**
2804	86	60	36
2903	86	62	36
3311	89	62	30
3373	91	63	26
3260	92	64	24
3305	93	64	22
3393←Best Heuristic value	**94**	**65**	**20**

Figure 4. Possible locations for a face in the image identified by the deformable template algorithm.

Unfortunately face recognition using the segment extracted by the face detection system yielded a recognition rate close to 0%[9]. This was because the model used for face recognition, principal component analysis, was sensitive to slight variations in shift, scale and rotation of a face image. Therefore to increase the suitability of the extracted segment for face recognition, a template matching system similar to the implemented two-tier face detection system was used for eye detection.

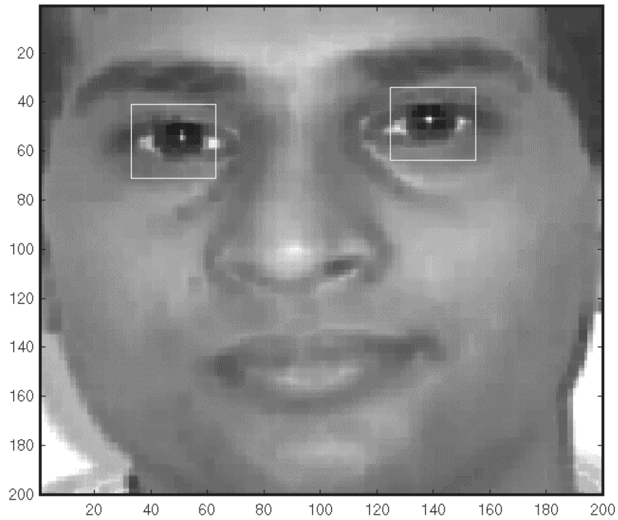

Figure 5. Successful eye detection of the pixel region identified by the face detection system. Once the accurate positions of the eyes are known, the test subject's face may be rotated and centred to increase its suitability for face recognition

The eyes of the test subject are approximately the same size and shape in all the extracted face images so the size of the template was not changed. Instead, to find the exact eye locations, the second tier of the eye detection system tested the correlation of the high heuristic (high 'eyeness') locations with *several* eye images which were of slightly different scales. This methodology proved to be the most suitable approach for eye detection.

Once eye locations were identified the test subject's face could then be rotated and centred based on the positions of the eyes in the extracted segment. Furthermore, once the exact positions of the eyes were calculated, the system was able to not only accurately extract a face image segment, but also eye, nose and mouth segments for recognition. The locations of the nose and mouth segments were estimated using the positions of the eye segments. Recognition could then be performed using all five extracted segments (left eye, right eye, nose, mouth and whole face segments).

3.0 FACE RECOGNITION

Recognition could have been attempted by directly comparing the raw pixel intensities of the extracted unknown image segments with known image segments. However, this technique would yield a very low recognition rate because all human face images are quite similar to one another. There is very little variability and high correlation between human faces because, after all, almost all of us have two eyes, a nose, mouth etc and have similar skin tones.

A typical extracted face used by this system would be a 100x100 image, i.e. a 10000-dimension vector. This face could also be regarded as point in 10000-dimension space, usually referred to as 'image space.'

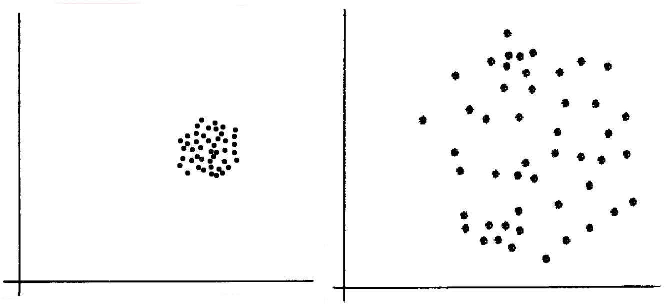

Figure 6. Faces in image space and face space. The data points in face space have a greater variability and therefore are more suitable for recognition.

To increase a face segment's suitability for recognition it is transformed from image space to 'face space.'[10] This transformation is based on principal component analysis, also known as the Karhunen-Loeve transform. Principal component analysis identifies variability between human faces, which may not be immediately obvious. It does not attempt to categorise faces using familiar geometrical differences, such as nose length or eyebrow width. Instead, a set of human faces is analysed to determine which 'variables' account for the variance of faces. In face recognition, these variables are called eigenfaces because when plotted they display an eerie resemblance to human faces. Any face image can then be described using these eigenfaces.

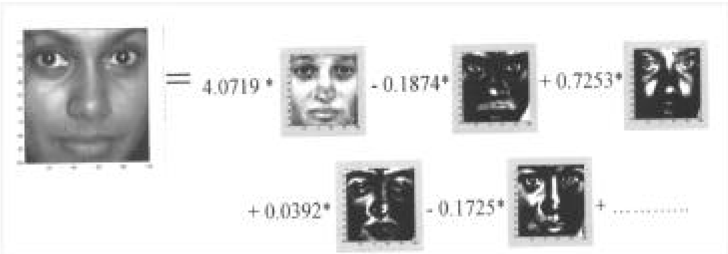

Figure 7. Graphical representation of the vector of a face in face space

When a face is projected from image space to face space, its face space vector consists of values corresponding to each eigenface. These eigenfaces are actually the eigenvectors of the covariance matrix of a set of mean subtracted face images (subtract the average face from each of the face images). The face images used should be a representative sample of the faces that the system would encounter. Since we are dealing with 10000-dimention vectors (i.e. face images), the resulting covariance matrix would be 10000x10000, and therefore computationally impossible for most modern computers. Therefore, the technique described by Turk and Pentland [11][12] was used to calculate the reduced covariance matrix's eigenvectors and the original covariance matrix's eigenvectors were deduced. Once the eigenvectors are calculated, for principal component analysis, we sort them according to their corresponding eigenvalue and take the required number of high eigenvalue eigenvectors. These eigenvectors account for the most variation of human faces. Therefore, Eigenface 1 describes more variation than Eigenface 2 and so on.

Figure 8. Eigenface 1 to Eigenface 9 displayed using a suitably scaled colour-map.

Since eye, nose and mouth segments were also extracted by the face detection system, these are also transformed into their respective vector spaces. An unknown face is recognized by transforming all its extracted segments into their respective vector spaces and finding the closest known individual to the transformed vectors.

4.0 RESULTS & CONCLUSION

The researcher gathered face images from 27 individuals to test the fully automated frontal view face detection and recognition system. Face images were intentionally taken under varying lighting conditions with the face being at different positions and scales in the image.

Successful results were obtained for automated face detection with a frontal view face detection rate of 100% being achieved using fully automated face detection. The complete fully automated face detection and recognition system with eye detection displayed a recognition rate of

73% on unknown face images. The researcher also implemented a manual face detection and automated recognition system to test recognition performance independent of the automated face detection and eye detection systems. This also yielded a recognition rate of 73%.

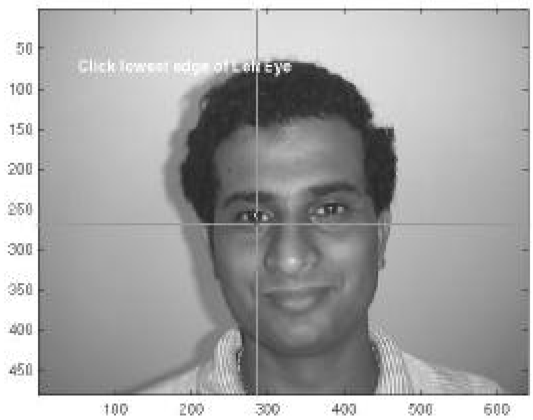

Figure 9. Manual face detection was used to test automated face recognition independent of automated face detection. A human operator was instructed to identify the exact face location in the image.

It may therefore be concluded that automated frontal view face detection has been very successful. The recognition rate of the entire system should be improved by enhancing principal component analysis face recognition. Since this study was limited to 27 test subjects, only 26 eigenfaces could be used for recognition. It is generally regarded that 40 eigenfaces can accurately represent a human face. Therefore by increasing the number of subjects in the study the recognition performance of the overall system will increase. This is in contrast to traditional neural network based techniques, where recognition accuracy would be adversely affected as the number of known subjects increases.

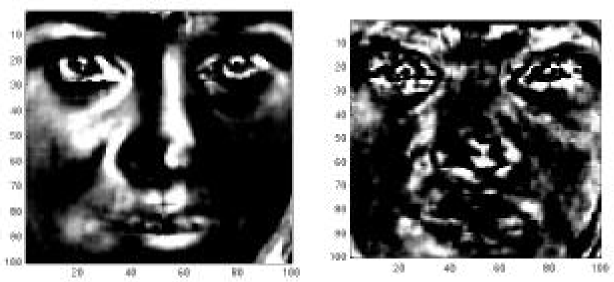

Figure 10. Eigenface 5 (left) and Eigenface 26 (right) displayed using a suitably scaled colour-map. It is apparent that Eigenface 5 is asymmetric and therefore was probably affected by lighting differences.

Further more, O' Toole at el. [10] showed that while large eigenvalue eigenfaces convey information regarding basic shape and structure it is the low eigenvalue eigenfaces that are useful for recognition. Therefore when many eigenfaces are used, not only would the 'image space' to 'face space' transfer become more one-to-one but the number of low eigenvalue eigenfaces would also increase dramatically, resulting in higher face recognition accuracy.

5.0 REFERENCES

[1] Sung, K. and Poggio, T. "Example-based learning for view-based human face detection." In Proceedings from Image Understanding Workshop, Monterey, CA, 1994

[2] Rowley, H., Baluja, S. and Kanade, T. “Neural Network-Based Face Detection.” Computer Vision and Pattern Recognition. 1996

[3] Kanade, T. “Picture processing by computer complex and recognition of human faces.” Technical report, Kyoto University, Dept. of Information Science. 1973

[4] Craw, I., Ellis, H., and Lishman, J.R. “Automatic extraction of face features.” Pattern Recognition Letters, 5:183-187, February. 1987

[5] Goldstein, A. J., Harmon, L. D., and Lesk, A. B. “Identification of human faces.” In Proc. IEEE, Vol. 59, page 748. 1971

[6] Kaya, Y. and Kobayashi, K. “A basic study on human face recognition.” In S. Watanabe, editor, Frontiers of Pattern Recognition, page 265. 1972

[7] Brunelli, R. and Poggio, T. “Face Recognition: Features versus Templates.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10):1042-1052. 1993

[8] Baron, R. J., ”Mechanisms of human facial recognition..” International Journal of Man Machine Studies, 15:137-178. 1981

[9] Balasuriya, L. S., "Frontal View Human Face Detection and Recognition," B.Sc.(Hons) Thesis, Department of Statistics and Computer Science, University of Colombo. 2000.

[10] O'Toole, A.J., Abdi, H., Deffenbacher, K.A., and Valentin, D. "A low-dimensional representation of faces in the higher dimensions of the space," Journal of the Optical Society of America, A:10: 405-411. 1993

[11] Turk, M. and Pentland, A. "Eigenfaces for recognition," Journal of Cognitive Neuroscience, 3(1): 71-86. 1991

[12] Turk, M. and Pentland, A. "Face recognition using eigenfaces," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June Maui, Hawaii, p 586-591. 1991