A FACE RECOGNITION SYSTEM USING HOLISTIC REGIONAL FEATURE EXTRACTION

Lawrence W. Cychosz and Jun Zhang

Department of Electrical Engineering
University of Wisconsin, Milwaukee, WI 53201

ABSTRACT

Two new face recognition systems are proposed using auto-associative back propagation neural network feature extractors on facial regions in conjunction with key facial structure measurements. A third proposed system combines the first two systems using confidence measurements to select a best match. A Cottrell/Fleming face recognition network and a structural face data network are also implemented and evaluated. A training set of 60 images and a test set of 12 images, acquired under uncontrolled conditions, were used to evaluate system performances. The first two proposed systems correctly selected 67 percent of the training images when presented with the test image set. The third proposed system achieved a recognition rate of 75 percent. By comparison, the Cottrell/Fleming network and the structural data network achieved recognition rates of 25 percent and 8 percent, respectively.

1. INTRODUCTION AND BACKGROUND

Studies have shown [1] that humans are born with an attraction to human faces; indicating some pre-wired general knowledge of faces. Research into a medical condition called prosopagnosia [2] has shown that it is likely that at least part of the human face recognition system operates by isolating key facial objects and then combining that information to perform face recognition. The tasks of combining and recognizing appear to be performed in a localized region of the brain. When we look at a human face, we acquire various information about it, as a whole and in regions, then our brains assemble the information and perform recognition.

Much previous research into automatic face recognition has fallen short in its applicability to the real world. Methods often rely on highly controlled environments; such as ambient lighting and head orientation, or make other idealistic assumptions. This paper evaluates a unique approach which is somewhat tolerant of changes in facial expression, hair style, scale, orientation, and ambient lighting.

1.1 Assumptions

There are many aspects to the face recognition problem. This paper addresses just one of those aspects; the identification of a face by selecting the best match from a set of known faces. It assumes that systems exist for locating faces in a complex background, for verifying that an image is a facial image, and for identifying key locations on a facial image; such as the corners of the eyes. These assumptions are valid since either solutions to these system limitations are available, or a significant level of successful research already exists [3][4][5].

1.2 Previous Work

Various research projects into face recognition [6][7] have utilized facial measurements, such as head width and the length of the nose. Most of these face recognition systems, however, have limited applicability outside of the controlled environment of the laboratory. Lim, Sim, and Oh [7], for instance, extracted seventeen structural face measurements from each image, granulized the data using fuzzy classifications, and then used a neural network to train their face recognition system. They achieved a 100 percent recognition rate using ten faces. Limitations of the system include the small size of their training set and a need for a highly controlled image acquisition environment.

Cottrell and Fleming [8] used an auto-associative back propagation feature extraction approach to create a face recognition system. They scanned sixty-four facial images and thirteen non-facial images at a resolution of 512x512 pixels and re-sampled them at a of resolution of 64x64 pixels. They then used a back propagation network as a feature extractor and a second neural network as a classifier. They found their system to have a recognition accuracy of 100 percent. Limitations of their approach include a low number of subjects (eleven) and a fairly large number of pictures per subject (an average of five).

Additional face recognition systems include WISARD [2], a general object recognition system based on statistical methods, and Photobook [9], a system which decomposes facial images into "eigenfaces", establishes all faces as linear combinations of the "most unique" images, and then performs comparisons based on the resulting combination vectors. These and other face recognition systems have strengths and limitations.

2. METHODOLOGY

2.1 Image Selection and Preparation

The training set for this research consisted of photographs of 60 people. The test set consisted of unique photographs of 12 of the people from the training set. All images were acquired from family photo albums and magazines. The face width was measured and the approximate scanning density required to acquire 300 pixels was calculated. This step performs a rough compensation for scale variations between pictures. A more exact compensation will be performed in a later step.

Next, the facial region was scanned using a 24-bit color scanner. The image was rotated so that the top of the lower eye lids were aligned horizontally; performing a gross adjustment for orientation. The pictures were cropped to include only the rectangular facial region; limited in width by the start of the ear and in height by the top of the eye brow (to eliminate contribution of the hair region). The bottom of the region was limited by the chin. If the chin was not visible due to a heavy beard, for instance, the location was approximated. This cropping step was added to prevent the undue influence of background images and of people's hair style on the systems under test. The image was then equalized. Since the pictures were acquired under uncontrolled lighting, this equalization step was added to compensate for some lighting and exposure effects.

2.2 Face Data Extraction

The next step was to identify specific coordinates and image regions. Figure 1 presents a representation of a face and the extracted data. Several coordinates were identified; including; inner corner of left eye, inner corner of right eye, bottom middle of lower lip, and bottom middle of chin.

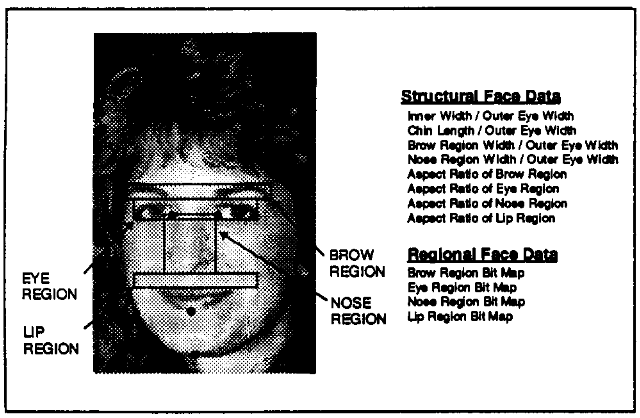

Figure 1: Manual Extraction of Face Information

Four facial regions were isolated and their width and height were noted. The images were re-sampled to compensate for scaling differences and were converted to 256 levels of gray. The eye, eye brow, and lip regions were re-sampled at 10x40 and the nose region was re-sampled at 20x20. The mouth region was not used since its shape can change dramatically.

Once recorded, structural face measurements were made scale-invariant by dividing the distance values by the distance between the outer eyes. Aspect ratios of the various regions were already scale-invariant. Values were then normalized by distributing them across the range of zero to one.

2.3 Holistic Regional Feature Extraction

Cottrell et al. [8][10] proposed and demonstrated an auto-associative neural network which performed image compression and whole face feature (holon) extraction. The algorithms proposed in this paper expand upon their work by using the known structure of the human face. Holistic features are extracted from the eye, nose, lip, and brow regions and, unlike prosopagnosia patients, the system combines the information to achieve recognition.

The four regional holistic feature extractors were created in a similar manner. The raw image files were normalized to values between zero and one. Each regional feature extraction neural network had 400 input nodes, 400 output nodes, and one hidden layer.

The combination of nodes per hidden layer, learning rate, and momentum were derived independently for each of the feature extraction networks using PRCPTRON, a custom back propagation neural network simulator [11]. The goal was to minimize the number of nodes required in the system's hidden layer so as to limit the development of grandmother cells. The maximum number of iterations was set to 25,000. The final number of nodes in the hidden layer was the minimum number of nodes required to achieve 100 percent successful recognition of the training patterns. Table 1 presents the final network configurations.

	Brow	Eye	Nose	Lip
Learning Rate	0.2	0.4	0.4	0.4
Momentum	0.6	0.6	0.6	0.6
Nodes / Hidden Layer	15	17	14	15

Table 1: Feature Extractor Networks

2.4 Face Recognition Methods

The structural face data recognition network was motivated by various research [3][7][12] into utilizing facial feature measurements for face recognition. This implementation used the eight structural face data values identified in Figure 1. The combination of nodes per hidden layer, learning rate, and momentum were derived by following a methodology for selecting the system with the smallest average pattern error energy while minimizing the number of hidden layer nodes [11]. The final network configuration had eight inputs, eight outputs, a learning rate of 0.7, a momentum term of 0.6, and a single hidden layer containing seven nodes.

The Cottrell/Fleming network [8] extracted holistic features (holons) from facial images with an auto-associative back propagation network which had sixty nodes in the hidden layer, a learning rate of 0.4 and a momentum of 0.6. The second step in the algorithm was to train another back propagation network classifier to match holon patterns with specific unique binary patterns. Figure 2 presents a graphical representation of the Cottrell/Fleming network.

The first of the newly proposed face recognition solutions combines the extracted structural data and the output of the holistic regional feature extractors into a vector. Facial image matching is accomplished by identifying the vector from the training set which has the shortest Euclidean distance to the vector derived from a given test image.

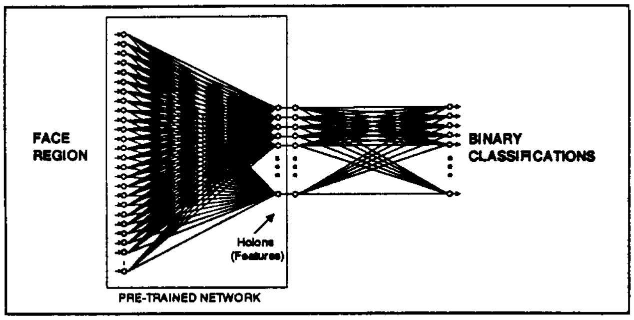

Figure 2: Cottrell/Fleming Classification Network

The second proposed solution uses the holistic regional feature extractors and the structural data as inputs to an auto-associative network. The intention of this approach is to force interactions between the different holistic features and the structural data. Matching is accomplished by identifying the training set image vector with the shortest Euclidean distance to a given test image's output vector. Figure 3 presents a graphical representation of the Auto-associative Combination Recognition Network.

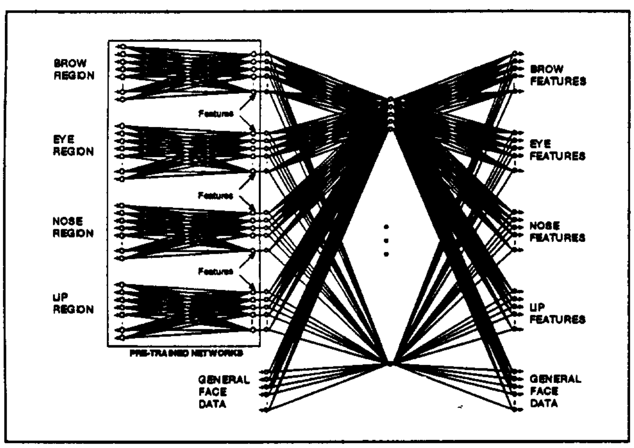

Figure 3: Auto-associative Combination Recognition Network

The third new face recognition solution combines the first two proposed solutions through a confidence mechanism. Each of the two previous systems calculates the Euclidean distance between the given test image feature vector and the selected training image's feature vector. It is postulated that a shorter distance between a test image output vector and the selected training set feature vector implies a higher level of confidence in the selection. The image with the smallest associated Euclidean distance is then chosen as a match.

3. RESULTS

The recognition rates for the five evaluated approaches are presented in Table 2. The Combination Confidence network, the third of the three newly proposed systems, demonstrated the highest recognition rate; 75 percent.

	Rate
Structural Face Data	8 %
Cottrell/Fleming Classifier	25 %
Regional Features and Structural Data	67 %
Auto-associative Comb Recognition	67 %
Combination Confidence	75 %

Table 2: Face Recognition System Recognition Rates

Two sample successful face recognition sets are presented in Figure 4.

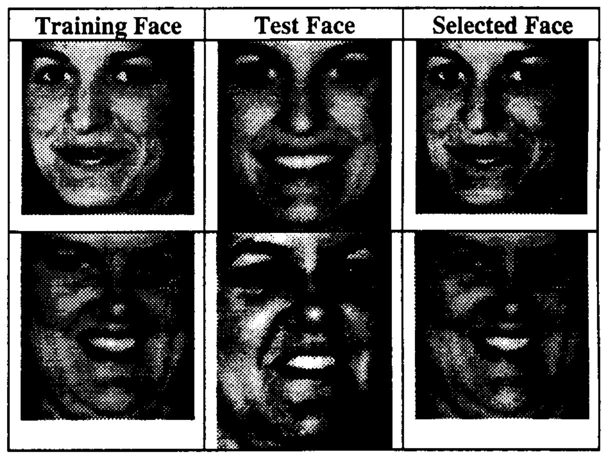

Figure 4: Sample Training, Test, and Selected Faces

4. DISCUSSION/CONCLUSIONS

The objective of this research was to create a face recognition system which would be somewhat tolerant of changes in facial expression, hair style, scale, orientation, and lighting. Three newly proposed systems were evaluated against two existing approaches. Each of the three new systems performed significantly better than either of the two previously existing approaches; the structural face data recognition network and the Cottrell/Fleming network.

There are a number of possible reasons why the two previously existing solutions performed at a much lower level than published. In the case of the structural data network, the eight selected measurements may not contain enough information to enable better performance; seventeen values were used my Lim, Sim, and Oh. Also, the orientation of a subject's head was highly controlled in previous work where it was only loosely controlled in this research. Finally, it is possible that using a selection set of consisting of sixty subjects instead of the original ten caused the drop in performance.

In the case of the Cottrell/Fleming network, this implementation used one view of each subject's face in the training set. Cottrell and Fleming used an average of six. Another factor is that Cottrell and Fleming selected best matching images from a training set consisting of only eleven individuals. Sixty images were used in this implementation.

The first of the new systems, the Regional Features and Structural Data network, performed significantly better than the Cottrell/Fleming network. One theory is that a holistic feature extractor focused on a very large area is not as good at extracting relevant features as are multiple regional feature extractors.

The second new system, the Auto-associative Combination Recognition network, was created with the intention of improving performance by forcing interaction between the various feature extractors. In evaluating performance, there was no difference in the gross performance of the Auto-associative Combination Recognition network and that of the Regional Features and Structural Data network. There was a difference between the actual selections, though. Of the eight correct matches of the two networks, six of the selections were the same.

The third system, the Combination Confidence network, combines the first two systems using a confidence measurement - improving system performances further.

Even though the Combination Confidence network achieved a correct recognition rate of 75 percent, it needs to have much better performance before it can be implemented in the real world. One means of improving the system might be to combine the features extracted from the whole image with those extracted from the regional images and the structural data. To enhance the system's ability to accommodate head rotations, another enhancement might be to use multiple views of a person in the training set. Three views consisting of a three-quarter view from the left, a three-quarter view from the right, and a straight on view would probably be sufficient. Another modification might be to include a capability marking poorly matching test images as "unknown". By requiring a minimum acceptable pattern error energy, poor matches could be identified.

REFERENCES

[1] C. Goren, M. Sarty and R. Wu, "Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants," Pediatrics, Vol. 56, pp. 544-549, 1975.

[2] H. Ellis and A. Young, "Are Faces Special?," In A. Young and H. Ellis (Eds.), Handbook of Research on Face Processing, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, pp. 1-19, 1989.

[3] R. Brunelli and T. Poggio, "Face Recognition through Geometrical Features," Proceedings of the Second European Conference on Computer Vision, pp. 792-800, Santa Margherita Ligure, Italy, May 1992.

[4] T. Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces," PhD Thesis, Department of Information Science, Kyoto University, 1973.

[5] Y. Guangzheng and T. Huang, "Human Face Detection in a Complex Background," Pattern Recogniton, 27, pp. 53-63, January 1994.

[6] G. Gordon, "Face Recognition Based on Depth and Curve Features," Proceedings of 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, pp. 808-810, June 1992.

[7] K. Lim, Y. Sim and K. Oh, "A Face Recognition System Using Fuzzy Logic and Artificial Neural Network," 1992 IEEE International Conference on Fuzzy Systems, San Diego, CA, pp. 1063-1069, March 1992.

[8] G. Cottrell and M. Fleming, "Face Recognition Using Unsupervised Feature Extraction," Proceedings of 1990 International Neural Network Conference, Paris, France, pp. 322-325, July 1990.

[9] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, HI, 1991.

[10] G. Cottrell, P. Munro, and D. Zipser, "Learning Internal Representations of Gray Scale Images: An Example of Extension Programming," Proceedings of the Ninth Annual Cognitive Science Society Conference, Seattle, WA, 1987.

[11] L. Cychosz, "A Face Recognition System Using Holistic Regional Feature Extraction," Master of Science Thesis, University of Wisconsin - Milwaukee, 1994.

[12] M. Kamel, H. Shen, A. Wong, and R. Campeanu, "System for the Recognition of Human Faces," IBM Systems Journal, Vol. 32, pp. 307-320, 1993.