Poster at the *IEEE Conference on Computer Vision and Pattern Recognition*, Boulder, CO, June 23-25, 1999.

A Probabilistic Framework for Embedded Face and Facial Expression Recognition

Antonio J. Colmenarez
www.ifp.uiuc.edu/∼antonio
Beckman Institute
University of Illinois
Urbana, IL

Brendan J. Frey
www.cs.toronto.edu/∼frey
Computer Science
University of Waterloo
Waterloo, ON

Thomas S. Huang
www.ifp.uiuc.edu
Beckman Institute
University of Illinois
Urbana, IL

Abstract

We present a Bayesian recognition framework in which a model of the whole face is enhanced by models of facial feature position and appearances. Face recognition and facial expression recognition are carried out using maximum likelihood decisions. The algorithm finds the model and facial expression that maximizes the likelihood of a test image. In this framework, facial appearance matching is improved by facial expression matching. Also, changes in facial features due to expressions are used together with facial deformation patterns to jointly perform expression recognition.

In our current implementation, the face is divided into 9 facial features grouped in 4 regions which are detected and tracked automatically in video segments. The feature images are modeled using Gaussian distributions on a principal component sub-space. The training procedure is supervised: we use video segments of people in which the facial expressions have been segmented and labeled by hand. We report results on face and facial expression recognition using a video database of 18 people and 6 expressions.

1. Introduction

In recent years, there has been a great deal of research on face recognition and facial expression recognition, but these two topics have been treated independently. The goal in most face recognition approaches is to find a similarity measure invariant to illumination changes, head pose and facial expressions, so that images of faces can be successfully matched in spite of these sources of variation. On the other hand, the goal of expression recognition is to find a model for non-rigid patterns of facial expression, so that expressions can be classified in spite of a wide range of variation.

To deal with variations due to facial expressions, several approaches have been studied. Elastic graphs (e.g. "Dynamic Link Architecture" [1, 2]) have been used to match images while allowing for non-rigid deformation . Another interesting approach uses a parametric model to deform faces so that the comparison of their appearance is carried out after the pose and facial expression is somewhat compensated [3, 4]. In a face recognition system based on a single frontal view, [5] uses the position of the facial features to estimate the head pose and normalize feature templates accordingly. In another approach for person identification based on the analysis of the spatio-temporal patterns of the lips [6], it has been found that combining shape information with intensity information improves recognition accuracy. Although there has been a great deal of research on the subject of facial motion analysis from image sequences, face recognition research has been focussed on still images. While there are several techniques somewhat invariant to variation in facial expression, only a few actually use the structure of facial expression deformation and none use these deformation patterns as part of the similarity measure for face recognition.

Most research on facial expression recognition is based only on non-rigid facial deformation patterns. Appearance variations due to facial expressions that are not well described by these motion fields are ignored. For example, these techniques use optical flow [7, 8], 2-D graphical models ("Potential Nets") [9] and local parametric models [10, 11].

In this paper, we propose a probabilistic framework in which face models jointly capture information about facial appearance and expression patterns so that recognition of faces and facial expressions are carried at the same time. In this framework, face and facial expression recognition cooperate so that the

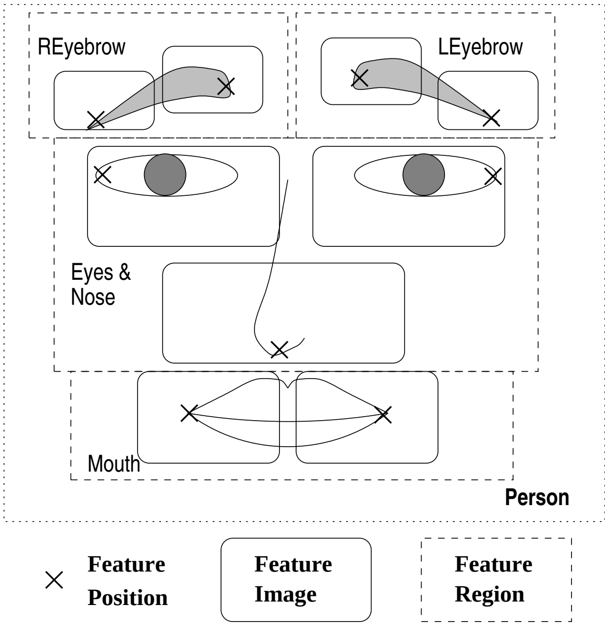

Figure 1. Scheme of the Facial Features and Regions.

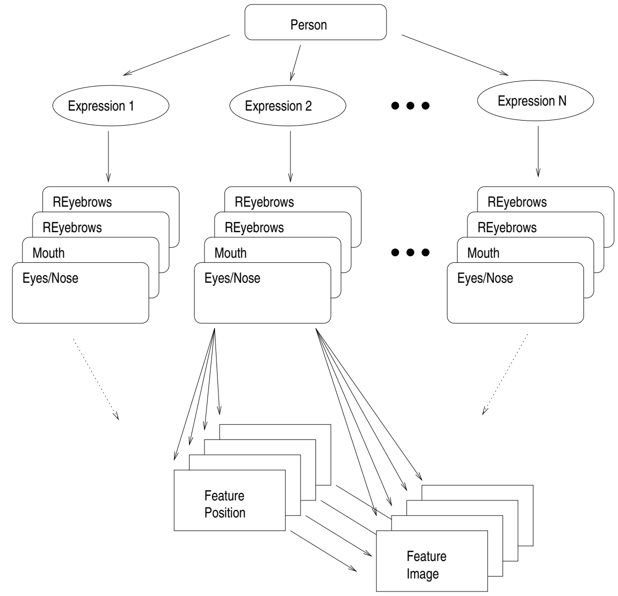

Figure 2. Probability Network for Face and Facial Expression Recognition.

similarity measure used for face recognition benefits from facial expression modeling. Conversely, expression recognition is improved by facial appearance modeling.

2. Modeling Facial Appearance and Expressions

Consider the face recognition problem where $\rho \in \{1, 2, \ldots, P\}$ is the index to ρ-th person in a database of P people and $\mathbf{f}$ is the portion of the observed image used for face recognition. Face recognition is carried out in a maximum likelihood setup,

$$\rho^* = \arg \max_{\rho=1,\ldots,P} \mathbf{P}(\mathbf{f}|\rho), \quad (1)$$

by selecting the model that maximizes the likelihood probability of the observed image.

We use a hidden, discrete variable $\epsilon = 1, 2, \ldots, N$ to index the facial expressions and the likelihood probability of the image $\mathbf{f}$ given the identity class ρ is computed from

$$\mathbf{P}(\mathbf{f}|\rho) = \sum_{\epsilon=1}^{N} \mathbf{P}(\mathbf{f}|\epsilon, \rho)\mathbf{P}(\epsilon|\rho) \quad (2)$$

In the proposed framework, faces are modeled as a set of regions containing sub-sets of facial features. This model is constructed under the assumption that the facial features can be accurately located. The appearance of each facial feature is provided by the image sub-window located around its position and the feature position is normalized with respect to the outer eye corners. Figure 1 illustrates the four facial regions and the nine facial features used in this implementation.

We assume the facial feature regions $\{\mathbf{r}_i \quad i = 1, 2, \ldots, R\}$ to be independent for given person and facial expression, and compute the likelihood probability of the observed image from:

$$\mathbf{P}(\mathbf{f}|\epsilon, \rho) = \prod_{k=1}^{M} \mathbf{P}(\mathbf{r}_k|\epsilon, \rho) \quad (3)$$

Finally, we compute the likelihood probability $\mathbf{P}(\mathbf{r}_k|\epsilon, \rho)$ of a region based on the position $\mathbf{x}_{ki}$ and appearance $\mathbf{v}_{ki}$ of the its R facial features as:

$$\begin{aligned} \mathbf{P}(\mathbf{r}_k|\epsilon, \rho) = \\ \mathbf{P}(\mathbf{v}_{k1}, \ldots, \mathbf{v}_{kR}|\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kR}, \epsilon, \rho) \\ \mathbf{P}(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kR}|\epsilon, \rho) \end{aligned} \quad (4)$$

Figure 2 illustrates schematically the proposed face model. Note that we used ovals to represent hidden states, round corner rectangles to represent the regions, and sharp corner rectangles for observed data.

We model the position of the facial features in a region $\mathbf{P}(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kF}|\epsilon, \rho)$ jointly with a multi-dimensional Gaussian distribution using a full-covariance matrix and mean vector estimated from the examples of each class. We model the appearance of each facial feature in a region independently

$$\mathbf{P}(\mathbf{v}_{k1}, \ldots, \mathbf{v}_{kF} | \mathbf{x}_{k1}, \ldots, \mathbf{x}_{kF}, \epsilon, \rho) = \prod_{i=1}^{F} \mathbf{P}(\mathbf{v}_{ki} | \mathbf{x}_{ki}, \epsilon, \rho). \quad (5)$$

We model the appearance of each facial feature with a multi-dimensional Gaussian distribution applied over the principal component subspace with p dimensions. Let $\mathbf{x} \in \Re^d$, be a d-dimensional random vector with some distribution that we are to model. We use a set of training samples of the class in question to estimate the mean $\bar{\mathbf{x}}$, and the covariance matrix Ω of the class. Using singular value decomposition, we obtain the diagonal matrix Σ corresponding to the p largest eigenvalues of Ω, and the transformation matrix $\mathbf{T}$ containing the corresponding eigenvectors. So, the conditional probability of $\mathbf{x}$ for a given class is computed from

$$\mathbf{P}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left[-\frac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}\right] \quad (6)$$

where $\mathbf{y} = \mathbf{T}(\mathbf{x} - \bar{\mathbf{x}})$ is the projection of $\mathbf{x}$ onto the aforementioned p-dimensional subspace.

3. Experiments and Results

In this section, we describe in detail the results of the experiments carried out to evaluate the performance of the proposed algorithm for embedded face and facial expression recognition. We first describe the video database we used for this purpose.

3.1 Face Video Database

We have tested our face and facial expression recognition algorithm using images from a video database of head-and-shoulder scenes. This database consists of 54 video segments containing 18 people with 3 distinct video segment each.

The first segment, intended mainly for expression recognition, consists of three repetitions of seven facial expressions in the following sequence: neutral, sadness, neutral, happiness, neutral, surprise, neutral, disgust, neutral, anger, neutral. The neutral expression was held for approximately 2 seconds to separate the others that lasted approximately 3 seconds each for a grand total of 2600 frames.

The second set of video segments consist of repetitions of facial gestures while showing three facial expressions: neutral expression, happiness and anger.

Figure 3. Sample frames from our face video database and the results of the facial feature detection/tracking system.

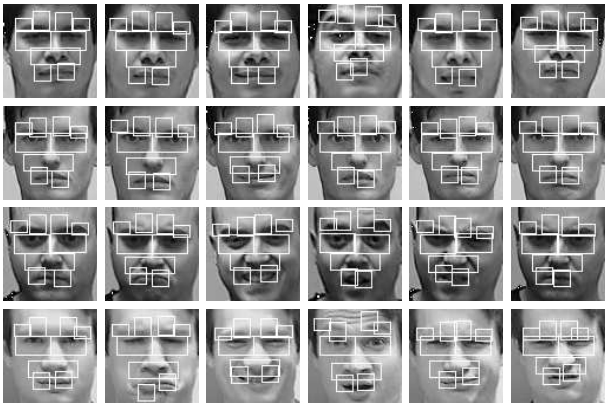

Figure 4. Examples of the facial feature windows and of the six facial expressions.

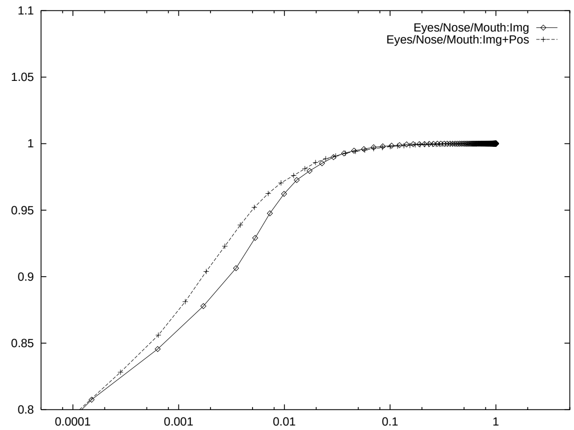

Figure 5. Face recognition performance using the images of the eyes, nose and mouth. The x-axis represents the false recognition rate and the y-axis the correct recognition rate.

The facial gestures include head nodding, head shaking, eye winking, and others. The total length of each of these segments is 2020 frames.

The last set of videos consist of 1350 frames of more relaxed gestures, expressions, and head motion intended for face tracking and recognition under extreme facial expressions.

All these videos were digitized in synchronism with a stimulus sequence (text commands on a laptop computer display) by detecting a high contrast change pattern that was displayed as part of the stimulus sequence several seconds in advance and recorded at the beginning of each segment. This pattern was easily detected by thresholding the frame difference, and the extra frames at the beginning of the sequence were skipped. We compressed and stored the videos in MPEG1 format, 320×240 color pixels, at 30 frames per seconds at an approximate rate of 1 Mbits/sec.

Figure 3 shows several example frames of these video segments and the result of our facial feature detection and tracking system [12, 13, 14]. The nine facial features detected are the outer eye corners, the eyebrow corners, the nostrils, and the mouth corners.

3.2 Face and Facial Expression Recognition

We trained one model for each person using the first two thirds of the frames of the first set of video segments in our video database and left the rest of the frames for testing our face and facial expression recognition algorithm.

Each model consists of four regions of features as illustrated in Figure 1 and 6 facial expressions: neutral, sadness, happiness, surprise, disgust and anger. These facial expressions were labeled by hand in the training video segments. Figure 4 shows examples of these six facial expressions and of the image windows of the facial features (e.g. $\mathbf{v}_{ki}$).

Face and facial expression recognition was carried out using maximum likelihood decisions. We approximate the likelihood of the face (2) with that of the facial expression that is most likely. So, the face recognition algorithm selects the person's model and the facial expression that maximizes the likelihood of the test image. In a person dependent scheme, the expression recognition algorithm simply selects the facial expression with maximum likelihood.

3.3 Face Recognition Results

We considered as a baseline for comparison, the face recognition performance using only the feature images in the regions of the eyes/nose and the mouth. Most other face recognition systems focus on these regions. Figure 5 shows the improvement in face recognition performance obtained by including feature positions in the similarity measurement as in (4).

Figure 6 shows the face recognition performance obtained using the feature images and positions in all facial regions. Note the improvement in performance obtained by including the eyebrows.

3.4 Facial Expression Recognition Results

Table 1 shows the facial expression recognition performance using only facial feature positions. Table 2

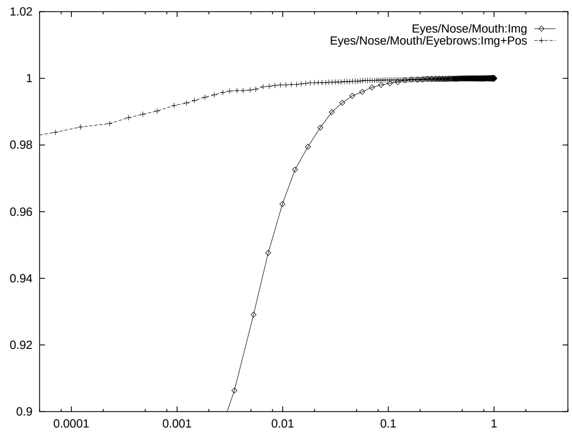

Figure 6. Improvement in face recognition performance using the eyebrow regions. The x-axis represents the false recognition rate and the y-axis the correct recognition rate.

shows the performance improvement obtained by using the facial feature positions and images. It interesting that using the feature positions *alone* gives reasonably good performance, although including the feature images significantly improves performance.

We also measure the expressiveness of the facial feature regions by comparing the recognition performance of each of the facial expressions independently. Figure 7 shows these comparisons.

4. Concluding Remarks

We have presented a Bayesian framework in which modeling facial feature appearance and modeling facial feature positions cooperate to improve the performance of face and facial expression recognition. We have shown that the use of facial feature appearances and positions in the similarity measure improves face recognition performance. Additionally, we have shown significant improvement in face recognition by considering the facial region of the eyebrows. We have also shown that expression recognition performance is improved by jointly modeling changes in facial feature appearance and position due to facial expression.

We are currently extending this framework to account for the temporal dependencies between facial expressions (collaboration with Zoubin Ghahramani, Gatsby Computational Neuroscience Unit). The hidden variable ϵ used to index the expression in (2) is replaced by a hidden state in a HMM. Face and facial expression recognition from video segments is then carried out using either the Viterbi algorithm or the

Neutral

Sadness

Happiness

Surprise

Disgust

Anger

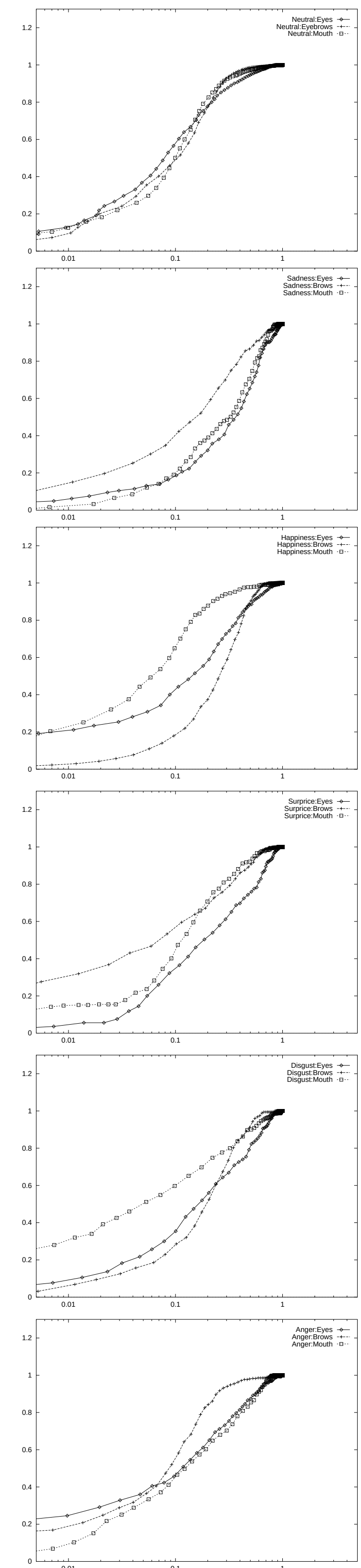

Figure 7. Expression recognition performance of each facial expression.

	Neutral	Sadness	Happiness	Surprise	Disgust	Anger
Neutral	4457	85	65	55	67	50
Sadness	908	501	29	15	43	45
Happiness	863	41	810	21	37	23
Surprise	626	14	21	612	25	17
Disgust	827	54	41	30	541	107
Anger	695	69	34	25	119	554
Error (%)	46.7	34.4	19.0	19.2	34.9	30.4

Table 1. Expression recognition performance using only feature positions

	Neutral	Sadness	Happiness	Surprise	Disgust	Anger
Neutral	7049	64	51	33	60	28
Sadness	337	633	3	6	44	47
Happiness	292	16	932	9	6	3
Surprise	125	2	6	700	3	0
Disgust	301	21	0	5	661	54
Anger	272	28	8	5	58	664
Error (%)	15.8	17.2	6.8	7.6	20.5	16.5

Table 2. Expression recognition performance using feature positions and images

forward-backward algorithm in a maximum likelihood setup.

Acknowledgments

This work was supported by Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0003. Brendan Frey was a Beckman Fellow at the University of Illinois at Urbana-Champaign when this research was conducted.

References

[1] L. Wiskott, J. M. Fellous, N. Krger, and C. Malsburg, “Face recognition and gender determination,” in *International Conference on Automatic Face and Gesture Recognition*, 1995.

[2] A. Tefas, C. Kotropoulos, and I. Pitas, “Variants of dynamic link architecture based on mathematical morphology for frontal face authentication,” in *CVPR*, 1998.

[3] A. Lanitis, C. Taylor, and T. Cootes, “A unified approach to coding and interpreting face images,” in *Proc. IEEE Conf. Computer Vision*, 1995.

[4] A. Lanitis, C. Taylor, and T. Cootes, “Automatic identification of human faces using flexible appearance models,” in *Procs. Of the 5th Brithish Machine Vision Conference*, 1994.

[5] K.-M. Lam and H. Yam, “An analytic-to-holistic approach for face recognition based on a single frontal view,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 673–686, July 1998.

[6] J. Luettin, N. Thacker, and S. W. Beet, “Learning to recognize talking faces,” in *Procs. of Int. Conf. On Pattern Recognition*, 1996.

[7] Y. Yacoob and L. S. Davis, “Recognizing human facial expressions from long image sequences using optical flow,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 636–642, June 1996.

[8] Y. Yacoob and L. Davis, “Computing spatio-temporal representations of human faces,” in *IEEE International Conference on Computer Vision*, (Cambridge, MA), IEEE Computer Society Press, June 1995, pp. 70–75.

[9] K. Matsuno, C.-W. Lee, S. Kimura, and S. Tsuji, “Automatic recognition of human facial expres-

sions," in *IEEE International Conference on Computer Vision*, (Cambridge, MA), IEEE Computer Society Press, June 1995, pp. 352–358.

[10] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motions," Xerox Palo Alto Research Center, Tech. Rep. CS-TR-3401, Jan. 1995.

[11] M. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *Proc. IEEE Int. Conf. Computer Vision*, 1995.

[12] A. Colmenarez, B. Frey, and T. Huang, "A real-time system for tracking multiple faces and facial features," in *Submitted to CVPR*, 1998.

[13] A. J. Colmenarez and T. S. Huang, "Pattern detection with information-based maximum discrimination and error bootstrapping," in *Proc. ICPR*, 1998.

[14] A. J. Colmenarez and T. S. Huang, "Face detection with information-based maximum discrimination," in *Proc. CVPR*, 1997.