Manifold analysis of facial gestures for face recognition

Douglas Fidaleo
Computer Vision and Robotics Research
Laboratory
University of California, San Diego
La Jolla, California USA
dfidaleo@ucsd.edu

Mohan Trivedi
Computer Vision and Robotics Research
Laboratory
University of California, San Diego
La Jolla, California USA
mtrivedi@ucsd.edu

ABSTRACT

The particulars of facial gestures are frequently used to qualitatively define and characterize faces. It is not merely the skin motion induced by such gestures, but the appearance of the skin changes that provides this information. For gestures and their appearance to be utilized as a biometric, it is critical that a robust model be established. For this purpose we are exploring gesture manifolds.

This paper describes work underway toward evaluation of the manifold representation of facial gestures both as a biometric, and as a means to extract biometric information. Details of the current acquisition system are discussed with the motivating principles behind the device. Preliminary observations are presented to motivate manifold analysis followed by an exposition of the experiments underway that will be used to validate the model.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—*Computer vision*

Keywords

Facial gesture analysis, biometrics, facial dynamics, G-folds, appearance manifolds

1. INTRODUCTION

> "The little twitch comes and goes so fast it's easy to miss. It's in the corner of [his] face, and it betrays the seething anger hidden beneath his mask of calm. It's not even anger, really, that seethes inside there - but indignation, that criminals are going free while the San Francisco Police Department devotes itself to public relations."
>
> - Roger Ebert on *The Dead Pool*

Clint Eastwood fans will instantly recognize the description above as Dirty Harry, but, what is it about the description that gives it away? The character is clearly male, and affected emotionally by the crime in San Francisco and the laissez fair attitude of the city's police department. But that merely narrows it down to a few million people. The giveaway is the twitch. The characteristic squinting of the eye that we have come to associate with the Dirty Harry character.

Considering that the eye twitch is such a strong indicator of Clint Eastwood's character, it is apparent that this, and other such "characteristic" gestures are useful in recognizing individuals. In other words, gestures may be useful as a biometric.

It is not, however, the twitch alone that identifies his character. While a gesture such as an Dirty Harry eye twitch or a George Bush smirk clearly helps define the person's character, it is not a sufficient description; there are many people with eye twitches and even more with characteristic smirks. Recognition involves assembling information from a variety of physiological and behavioral observations of a subject. Gestures can, however, reduce the uncertainty present in other unimodal biometrics and enhance recognition.

The key problem (and focus of our work) is to find robust operators and computational framework for characterizing both the temporal and appearance characteristics of facial gestures. We are exploring a model of facial gestures based on appearance manifolds (G-folds) that encapsulates spatio-temporal appearance information of a facial gesture in a single structural model [6]. Properties such as asymmetry, gesture intensity, and gesture dynamics can be easily extracted using a controlled spatial decomposition of the face.

Gesture manifolds [6] are an interesting application of parametric appearance models [16] to the analysis of facial gestures. In [7] the face is partitioned into local co-articulation regions. Facial gestures analyzed in these regions exhibit a coherent correlation structure in the form of paths in PCA space parameterized by gesture intensity. G-folds are well suited to real-time gesture analysis on trained subjects and are applied in [6] to control of facial animation.

This work is also related to modular eigenspace approaches [19] where PCA subspaces are constructed at local face features. In G-folds however, partitions are tailored to gestures (dynamic features) rather than static facial features such as the nose and eyes. Eigenspaces are constructed using temporally varying images from a single subject, rather than static images across several subjects. Given the distinctive

WBMA'03, November 8, 2003, Berkeley, California, USA.
Copyright 2003 ACM 1-58113-779-6/03/00011 ...$5.00.

characteristics between the correlation curves of different subjects, it was suggested that the G-folds may be effective as biometric signatures [17].

The intent of this paper is not to *prove* the utility of the G-Folds representation, but to *motivate* its potential. The remainder of the paper will detail the experiments currently underway in the CVRR lab at UCSD to assess both the theoretical and practical ramifications of the G-fold model for use as a fundamental biometric and as an analysis tool for extracting biometric information. This paper will also detail the infrastructure developed that is necessary to carry out such experiments with confidence.

2. VIDEO-BASED DYNAMIC BIOMETRIC FEATURES: RELATED STUDIES

The idea that expressions contain useful information for recognition of faces is not new. Significant work has been performed to assess the effects of facial dynamics (vs. static faces) on subject memory and subsequently a subject's ability to recognize new faces, possibly from novel view angles or lighting conditions. Particularly, evidence suggests that facial dynamics enhances recognition accuracy under suboptimal viewing conditions [3].

Dynamics in the form of motion information is the focus of most expression based recognition work[21][14]. Early work by Basili showed a correlation between facial motion and the perception of various key emotions [1]. As indicated in [9] motion can also provide knowledge of subject gender.

While motion information is clearly part of the story and serves to normalize variations due to lighting, skin color, and other static facial variations, it loses significant information that occurs as a result of dynamic skin wrinkling and static appearance characteristics of the skin. These appearance variations provide important evidence towards subject identity, both in the spatial and temporal domain as indicated in O'Toole et. al. [18].

In an attempt to tease apart the roles of facial geometry (form) vs. motion in identification of individuals, Knappmeyer et. al. explored the use of animated 3D models of human faces. Results indicate a strong motion bias and provide evidence that form and motion are "integrated during recognition, rather than operating as independent cues." [12] This provides very strong motivation for a biometric feature set that combines temporal and static appearance information.

A similar conclusion can be drawn from their experiments showing that the strongest motion bias is found in the absence of facial texture (when the 3D geometry is rendered without skin texture from the acquired subject.) When texture is included, motion becomes less important in identifying the face. This speaks strongly to the fact that facial appearance plays an important role in face identification.

A very relevant approach to shape and texture unification called active appearance models has been applied to face interpretation[4][5]. Unfortunately, this model neglects the temporal component of appearance.

Video based person tracking, body modeling, movement and body tracking, face detection and recognition, and affect analysis have been important areas of research in our group. We have developed and deployed a series of systems for multi-person tracking [20][11], face capture and pose estimation [11][15], face recognition [10], and facial affect analysis and face modeling [15] [7] [6] [8]. A common thread running in all of these efforts is the dynamic analysis of the variations captured by video signals, which results in accurate, reliable, and robust algorithms.

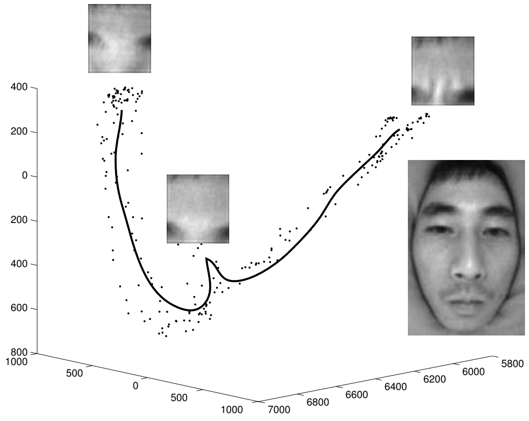

Figure 1: Projection of gesture data for the eyebrow raise and brow furrow gesture onto the first 3 principal components of X. The principal trajectory of each gesture actuation is shown in the solid curve.

3. G-FOLDS

This section will give a brief overview of the G-Folds representation, but the reader is referred to [6] for a detailed description.

The motivation for appearance manifold analysis is that images with higher correlation in image space will be closer in a reduced dimensionality PCA space [16]. Processes with significant correlation over small parameter variation will induce an appearance manifold that is a function of the parameter(s). In [6] the varied parameter is gesture intensity, and indeed, there is significant correlation between similar-intensity samples.

Figure 1 shows two gestures (FROWN and BROW FURROW) extracted from the center forehead region on a subject's face. The gestures are each actuated three times from a neutral starting point to full muscle contraction. The successive intensity images of the gestures are reduced to three dimensions using PCA and plotted in the reduced subspace. It can be seen that the gestures trace out coherent paths as the gesture progresses from neutral to maximum actuation. The induced gesture structures are referred to as *gesture manifolds* (G-Folds).

The set of discrete manifold samples of each gesture is further reduced to a 1-D continuous curve by a second PCA on the manifold samples for individual gestures and quadratic polynomial regression. The curve parameter varies with gesture intensity and hence the intensity of a new incoming sample is determined by projection onto the polynomial.

4. EXPERIMENTS

We are pushing forward on two biometric applications of this model. The first is a natural extension of [6] exploring G-folds as a tool to extract gesture intensity parameters. By analyzing gesture intensity, we have access to other information such as gesture asymmetry and temporal actuation

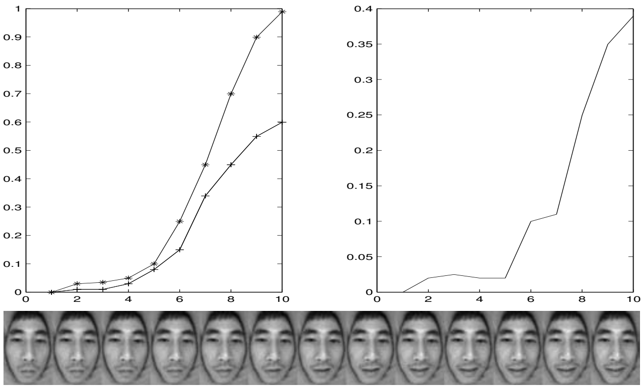

Figure 2: (left) Gesture intensity profiles for left and right smile gesture. (right) Actuation intensity difference between profiles.

patterns. On a second front, we are evaluating the manifold structure itself as an explicit biometric signature. The gesture manifolds encapsulate static and dynamic appearance information in a single concise model. We are pursuing closed-form and machine learning methods to compare the G-fold signatures.

4.1 G-folds as an analysis tool for biometric feature extraction

The G-fold representation was used in [6] for analysis of the intensity of a specific set of facial gestures and used to control facial animation sequences. There are several potential physiological and behavioral biometric features that can be extracted using such an intensity analysis tool. In the behavioral domain we are currently exploring:

Gesture activation profile The activation profile over time.

Gesture frequency The frequency of actuation of a given gesture, potentially co-occurring with other gestures.

Gesture context External factors that elicit the gesture.

Facial asymmetry Both static and dynamic asymmetries.

Of note, facial asymmetry was shown in [13] to contain relevant information for face recognition. Asymmetry in that work was characterized holistically (using the entire face) using a spatio-temporal metric. Asymmetry in our case can represented explicitly for particular gestures.

Figure 2 shows an illustrative example of the potential of G-folds for asymmetry. The left plot shows gesture intensity over time for the left and right side of the mouth during the smile gesture. The differences in activation levels over time are shown in the right plot. This difference plot characterizes the dynamic facial asymmetry as the smile gesture is actuated. Other similar secondary signatures are made available by gesture intensity analysis.

4.2 G-Folds as a biometric signature

The distinct structure of gesture manifolds suggests that they may function well as a biometric signature. However, similar to the lip-profile example of Brand et. al. [2] G-folds are difficult to classify as a behavioral or physiological biometric in isolation. Arguably, this distinction is blurred further in the case of G-Folds as dynamics are only implicit in the signature, in fact the representation effectively removes the temporal dependence.

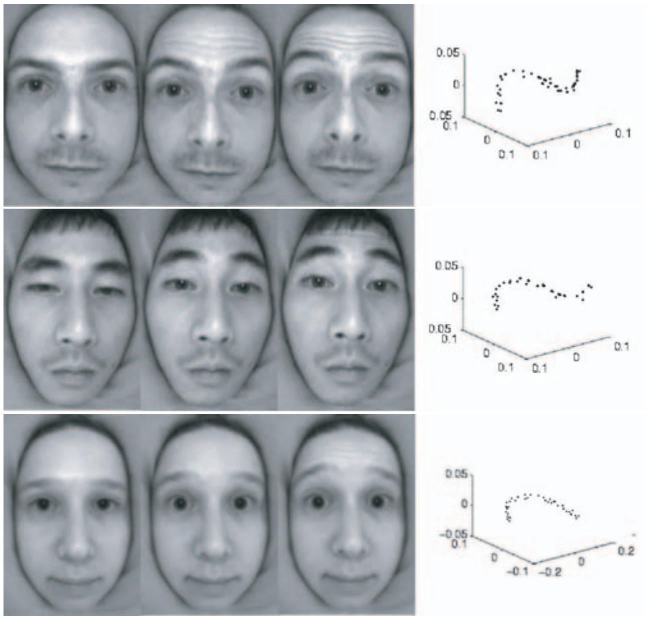

Figure 3: Illustrations of 3D manifolds for three subjects A(top), B (mid), and C (bottom) actuating the eyebrow raise gesture.

To motivate the G-fold representation, the reader is refered to Figure 3. This figure illustrates a set of manifolds for various facial gestures from three subjects. These manifolds share similar structure (they traverse S-like curves in appearance space), but there are also significant differences between the manifolds of all three subjects. Qualitatively comparing the two most similar manifolds, we see that the tail of subject A is shorter, and the upper dip is less shallow than subject B.

Figure 4 illustrates an independently observed actuation of the eyebrow raise gesture. Qualitative comparison of this observation to the manifolds in figure 3 shows greatest similarity to subject B.

While the evidence suggests that the appearance dynamics encapsulated in the G-fold model are important for face

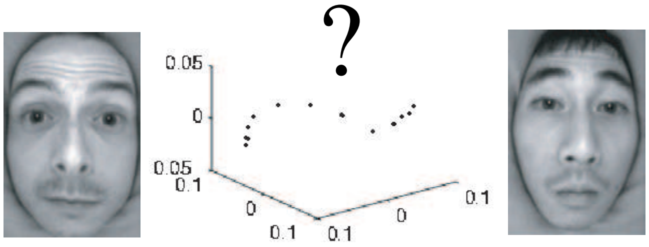

Figure 4: Which subject generated this manifold?

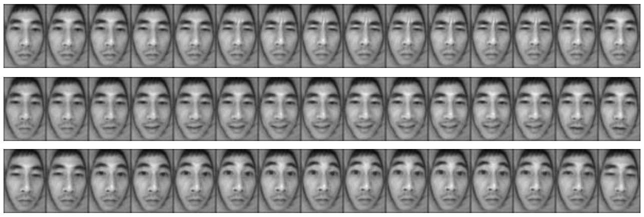

Figure 5: Test sequences for the furrow, smile, and eyebrow raise gestures.

recognition, there are three fundamental questions we hope to answer with the current experiments.

1. What is the raw recognition accuracy achieved with G-folds signatures alone?
2. Can G-folds be used to enhance existing biometrics?
3. How can we combine G-folds signatures with gesture intensity derived metrics for a combined physiological and behavioral biometric.

5. EXPERIMENTAL SETUP

5.1 Gestures

To make an initial assessment of G-Fold utility we are considering 3 gestures: Eyebrow raise, smile, and brow furrow shown in Figure 5. These gestures were selected primarily for their ease of actuation. Each gesture is actuated six times from neutral to maximum. To test the model under somewhat varying internal conditions (mood, fatigue, facial hair) full data sets are acquired from each subject in three different sessions over the course of one month.

5.2 Infrastructure

The requirements for G-Fold analysis are shared with most facial analysis systems. At a high level, we must decide upon the variations we are interested in uncovering, and normalize the remaining unwanted variations. For our experiments, we are primarily interested in the appearance changes due to facial gestures. Consequently, variations due to lighting and pose are removed. While there are several signal processing type methods for reducing these variations, each introduces an element of uncertainty. In the interest of uncertainty reduction, we have opted to construct a testbed that minimizes the potential for head pose variation and completely controls lighting.

The device consists of a closed wooden box with an opening for the subject to place her head (Figure 6). The opening is cushioned which serves to both comfortably retain the head position in a relatively fixed position and orientation, as well as match the contours of the face and thus block incoming light. Facial gestures and speech exhibit rapid variations that are blurred by standard (30Hz) video sensors. A high frequency (50Hz) camera is therefore mounted at the rear of the box, facing the subject to acquire images. A point light source with diffusion material is mounted above the camera to illuminate the face.

Recent trends towards multi-modal integration indicate the need for multiple sources of information to reduce the natural uncertainty present in a single source. Though gestures are the focus of this work, we recognize that gestures alone may be insufficient as a biometric. The acquisition system has therefore been outfitted with a microphone array and thermal imaging device for acquisition of speech and thermal signatures of gestures.

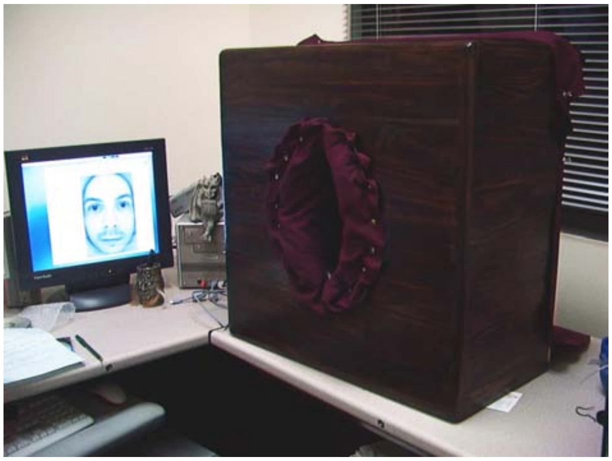

Figure 6: Acquisition device developed as part of our testbed for multimodal analysis of facial appearance, dynamics, and form.

The system consists of three basic pre-processing units: Sampling, Manifold Extraction, and Intensity Analysis. Two biometric feature extraction units: one for extracting behavioral biometric information from gesture intensity data, and another for physiological biometric manifold extraction. A final biometric feature composition unit serves to merge features and compare existing signatures. The data flow through these modules is shown in figure 7.

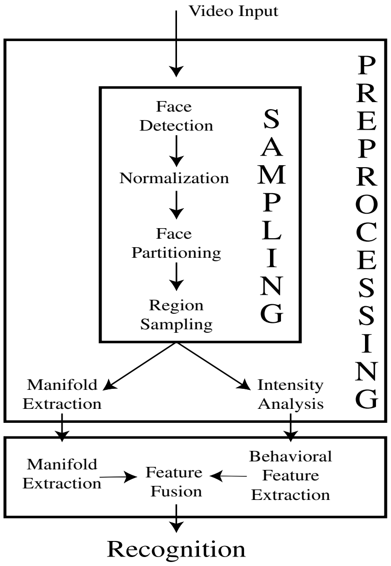

Figure 7: Experimental components

As G-folds are an appearance based model, small variations in alignment between samples can translate into large variations in the extracted manifolds. Face detection and tracking are therefore important components in normalizing the pose variations of expressive images. For face detection and tracking we are currently using a Gabor wavelet feature detector/tracker.

6. CONCLUSIONS

Distinctive facial gestures clearly assist in defining a person's character. The work underway in the CVRR lab at UCSD is evaluating the possibility that facial gestures contain sufficient person-specific information to *identify* a person. The G-folds representation developed at USC is a rich quantitative model of gesture appearance that we believe holds great promise for biometric signature extraction.

The initial experiments are focused on the gestures listed in the previous section, but manifold analysis is equally applicable to other constrained facial gestures. Following validation of the G-folds model, we intend to expand the work to include both thermal and aural signatures. The constructed testbed enables the combination of these modalities for more robust biometric operator construction. The set of analyzed gestures will also be extended.

The model has already been tested with great success for person-specific gesture analysis [6]. Though quantitative results are not currently available, our current observations indicate that G-folds are also excellent signatures for discriminating between subjects. We are excited about further pursuit of G-folds as biometric operators.

7. ACKNOWLEDGMENTS

We are thankful for the grant awarded by the Technical Support Working Group (TSWG) of the US Department of Defense which provided the primary sponsorship of the reported research. We are also thankful for the contributions and support of our colleagues from the UCSD Computer Vision and Research Laboratory.

Professor Ulrich Neumann of the University of Southern California deserves a special note of thanks and appreciation for suggesting to the author (D.F.) the possibility of G-Folds based biometric signatures and initial discussions of their application to face recognition.

8. REFERENCES

[1] J. N. Basili. Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology*, 4:373–379, 1978.

[2] J. D. Brand, J. S. D. Mason, and S. Colomb. Visual speech: A physiological or behavioural biometric? *Lecture Notes in Computer Science*, 2091, 2001.

[3] V. Bruce. Fleeting images of shade: Identifying people caught on video. *The Psychologist*, 11(7), July 1998.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407, 1998.

[5] G. Edwards, C. Taylor, and T. Cootes. Interpreting face images using active appearance models. In *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition*, pages 300–305, 1998.

[6] D. Fidaleo. *G-Folds: An appearance based model of facial gestures for performance driven facial animation*. PhD thesis, University of Southern California, 2003.

[7] D. Fidaleo and U. Neumann. Coart: Coarticulation region analysis for control of 2d faces. In *Computer Animation 2002 Proceedings*, pages 17–22, 2002.

[8] D. Fidaleo, J. Noh, T. Kim, R. Enciso, , and U. Neumann. Classification and volume morphing for performance-driven facial animation. In *International Conference on Digital and Computational Video*, 1999.

[9] H. Hill and A. Johnston. Categorizing sex and identity from the biological motion of faces. *Curr. Biol.*, 11:880–885, 2001.

[10] K. Huang and M. Trivedi. Streaming face recognition using multicamera video arrays. In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 4, pages 213–216, 2002.

[11] K. Huang and M. Trivedi. Video arrays for real-time tracking of person, head, and face in an intelligent room. *Machine Vision and Applications*, 14(2):103–111, June 2003.

[12] B. Knappmeyer, I. Thornton, and H. Buelthoff. Facial motion can determine facial identity. *Journal of Vision*, 1(3), 2001.

[13] Y. Liu, R. Weaver, K. Schmidt, N. Serban, and J. Cohn. Facial asymmetry: A new biometric. Technical Report CMU-RI-TR-01-23, Carnegie Mellon University, Pittsburgh, PA, August 2001.

[14] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions*, 74(10), 1991.

[15] J. McCall, S. Mallick, and M. Trivedi. Real-time driver affect analysis and tele-viewing system. In *Proceedings of IEEE Intelligent Vehicles Symposium*, pages 372–377, June 2003.

[16] S. Nayar, H. Murase, and S. Nene. Parametric appearance representation. In *Early Visual Learning.* Oxford University Press, February 1996.

[17] U. Neumann. Personal correspondence. University of Southern California, April 2003.

[18] A. O'Toole, D. Roark, and H. Abdi. Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6:261–266, 2002.

[19] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition*, 1994.

[20] T. Sogo, H. Ishiguro, and M. Trivedi. *N-Ocular Stereo for Real-Time Human Tracking*, chapter Panoramic Vision: Sensors, Theory, Applications, pages 359–375. Springer, New York, 2001.

[21] Y. Yacoob and L. Davis. Recognizing facial expressions by spatio-temporal analysis. In *12th International Conference on Computer Vision and Pattern Recognition*, 1994.