(Excerpt from the PhD Thesis entitled 'Using 3D Facial Motion for Biometric Identification', 11/2009)

PhD Thesis - Chapter 5

Quantification of 3D Facial Motion

Author: Lanthao Benedikt

Supervisors: Dr. David Marshall, Dr. Paul Rosin

April 11, 2010

Cardiff University - School of Computer Science

Address: Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom

Website: http://www.cs.cf.ac.uk

Thesis Abstract

In the current context of heightened security, expanded research effort is needed for exploring novel biometric modalities, while continuing to improve existing solutions. The purpose of this research is to investigate whether the 3D spatiotemporal dynamics of facial expressions can be used for person identification, and to compare the advantages of this novel approach to classic face recognition using static mugshots. The working hypotheses are formulated as follows:

> ***Hypothesis 1:*** *Human facial movements are stable over time and sufficiently distinctive across individuals to be used as a biometric identifier. However, there exists a hierarchy in the reproducibility and uniqueness of facial actions, such that some facial expressions are more suitable for person recognition compared to others.*
>
> ***Hypothesis 2:*** *There exists a motion signature associated with moving faces, which describes not only the behavioural traits of a person, but also reveals physical idiosyncrasies that becomes visible only when the face is in motion. This identity cue is independent from the physical information conveyed by static faces, such that the temporal dynamics of facial expressions contribute to improve identity perception.*

The purpose of the present chapter is to review existing methods to model the human head and its motions, in light of which the best approach is identified and implemented.

1 Survey of feature extraction methods

Research into the modeling of the human face and its motions over the last four decades have led to the development of many solutions that have been reported in the literature. The following section offers an overview of the major facial feature extraction strategies and algorithms.

1.1 How to quantify facial motion?

Facial motion produced by emotional expressions or speech is a complex sequence of facial muscle activations that can be described in different ways:

- **As an 'action unit profile':** this approach consists of using high-level feature descriptors for encoding facial movements.
  - An example of such method was developed in 1978 by Ekman and Friesen who proposed FACS (Facial Action Coding System), a system that describes any facial expression as a combination of Action Units (AU) and the temporal occurrences of these [1]. This method has since gained significant popularity among researchers in the field of facial expression recognition for quantifying emotional expressions [2–4].
  - Another example of high-level feature descriptors can be found in research on speech synthesis and lipreading, where facial movements can be described as combinations of speech-related action units (or lip shapes) commonly designated as visemes [5–7].
  - In practice, given a sequence of facial performance, an action unit recogniser is first applied, then classification algorithms such as the Levenshtein distance [8], the Pearson's correlation coefficients [3], Hidden Markov

Models [5], Gaussian Mixture Models [6], or wavelet features and SVM [4] can be used to compare action unit profiles.

- **Manifold subspace analysis:** the features are learned directly from the face data without using any intermediate feature descriptors [9]. Given that the human face complies to strict spatial constraints (e.g. large smooth surface, facial feature location, symmetry) [10] and deforms smoothly during expression [11], the high-dimensional raw face data can be described by a small set of parameters; such quantification and dimensionality reduction make it easier and computationally more efficient to analyse and compare facial performances. Thus, the facial dynamic extracted from a video input is represented as a time series in a low-dimensional subspace. Figure 1 shows an example of using manifold subspace analysis to extract facial motion features from video data. In the example below, 2D video data is being analysed. In the following sections of this report, we will further examine manifold analysis of three-dimensional video data.

- **Deforming surface classification:** a method for characterising transitions between curvature classes over time has been recently developed by a research group at the University of Edinburgh [12]. This can also be used for analysing dynamic 3D data because changes in local curvature can be considered as a feature set that encapsulates individual differences during facial expressions. This method is likely to be robust in the presence of reconstruction noise because curvature is computed over a local 3D patch [11].

- **Other techniques:** there also exist many novel techniques to encode dynamic facial features. For example, Pamudurthy et al. [13] proposed the Digital Image Skin Correlation method where dense correspondences at the skin pore

level is computed and used to compare images. Zhang et al. [14] established an identity signature based on the elastic characteristics of facial tissue by computing the strain pattern of the face in the closed and open jaw positions; this required the manual mark-up of about 400 corresponding points between the two images. These methods have not been so far fully validated and will not be considered in this study.

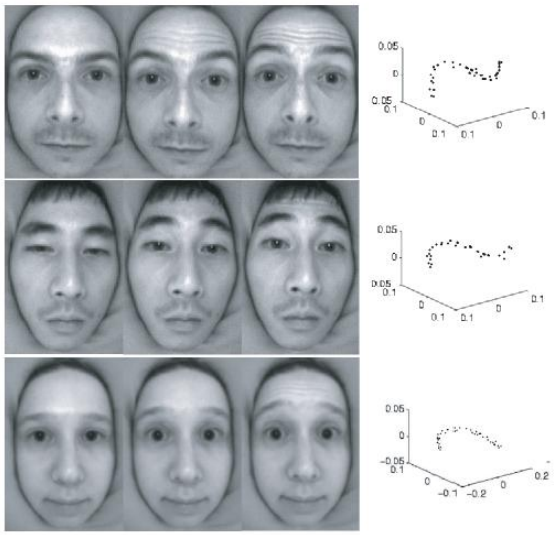

Figure 1: Experiment described in the work of Fidaleo et al. [9]. Manifold analysis is applied to extract facial motions from 2D video data.

At the moment, the most widely preferred method for quantifying facial motion is to use action unit profiles (FACS and viseme decoding). This approach is very useful for applications in facial expression recognition and lipreading, where there is a need to categorise and identify the content of the facial performance (i.e. what emotion is it? What is the person saying?).

However, for identity recognition, the facial performance is usually known. Therefore, it is arguable whether the use of intermediate feature descriptors is helpful, or

on the contrary, this would introduce an unnecessary complexity and an additional source of error due to the existing AU-recognisers that are not yet 100% reliable. Therefore, the manifold subspace analysis and the deforming surface classification appear to be better choices. The next section reviews a number of well-established manifold subspace analysis methods and discuss their strengths and weaknesses.

1.2 Feature extraction algorithms

Methods for representing facial forms and motions vary in how closely they model the anatomical structures of the human head and exploit the shape information:

- At one extreme are the *physically-based models* where the human face is treated as a biomechanical system including highly accurate details of the underlying facial anatomy such as bones, musculature and skin tissue. Examples of such models include the pioneering works of Waters [15], Essa [16] and in particular that of Terzopoulos [17] where the facial soft tissues are described as various layers of elastic spring meshes, the unit actions are simulated by forces and the deformations are determined by solving dynamic equations. Twenty years ago, this approach was considered very promising for modeling accurate facial deformations in realistic animations [16,18,19]. However, it has not ultimately enjoyed the predicted popularity because of the high computational cost associated with the level of fidelity required [20, 21], another limiting factor being the lack of anatomically accurate model, as reported by Sifakis et al. [22]. Despite these shortcomings, there have been several recent attempts to develop such anatomical face models [23–25].

- At the other end are methods such as the works of Yacoob and Davis [26], Mase and Pentland [27], DeCarlo and Metaxas [28] that use optical flow to identify

the direction of facial motions, utilising only very weak information of facial shapes. The main advantage of these methods is their simplicity, as they rely on very little prior knowledge of the studied object. However, these techniques can only recover motion orthogonal to the intensity gradients and generally use a small spatial area to determine each motion vector. As a consequence, they are very sensitive to noise and fall short of reliable tracking [11], and generally, they cannot capture rapid and subtle motions [18].

- An intermediate approach is the *active contour models* that treats the human face as a deformable surface without getting down to the details of the underlying facial anatomy. These models offer a compromise between accuracy and complexity to describe visually observable facial deformations. Facial key features (e.g. the lip contour, the eye corners, the nose tip) are tracked across the video sequence and their displacements are used to recognise facial motions. Early tracking methods include *snake*, an energy minimising spline guided by external constraint forces that pull it toward lines and edges [29,30]. A more elaborate method is the CANDIDE model that aims to fit a simple triangulated mesh defined by a set of facial features to the observed data using constrained optimisation [31]. Later, Cootes et al. proposed the more complex deformable models e.g. the ASM (Active Shape Model) for 2D shape analysis [32] and the AAM (Active Appearance Model) that combines both 2D shape and texture [33]. Another example of deformable model is the Morphable Model developed by Blanz and Vetter for 3D data analysis [34].

There exists another category of techniques for representing the human face that is usually referred to as the *appearance-based* methods, in contrast to the *model-based* methods described above. These include classic algorithms e.g. Eigenfaces (PCA) [35], Fisherfaces (LDA) [36], and ICA [37]. Although these are well-established

methods that have been successfully employed in many static face recognition problems [38–40], they are considered less sophisticated than model-based methods that are capable of learning constraints from the training data, which allows them in a later stage to fit to new shapes in unseen images (i.e. 'scene interpretation').

2 Feature extraction strategy

The contribution of different facial regions to perception of identity has attracted great interest within research in psychology [38, 41] where three face recognition strategies have been identified: holistic, feature-based and hybrid. Figure 2 shows the feature extraction algorithms that can be employed for each of these strategies.

2.1 Holistic Approach

The holistic approach considers the entire face for recognition. It has been demonstrated that high dimensional face images belong to a subspace of much lower dimension [10], thus the densely sampled face images are first projected into a low-dimensional subspace defined by a set of basis vectors, then the projection coefficients are used for comparing faces. The most common classification algorithms include the nearest-neighbor search [35] and the probabilistic Bayesian approach [42].

Popular linear methods include Eigenfaces that applies PCA to extract the most characteristic features [35], ICA that generalises Eigenfaces by considering higher-order statistics [37], and Fisherfaces that uses LDA to extract the most discriminant features across individuals [36]. Other methods such as SVM (Support Vector Machine) [43] and 3D Morphable Model [34] have also been considered for face recognition. The performance comparison between these different methods is still a subject of debate, as reported in the literature [39, 44–46].

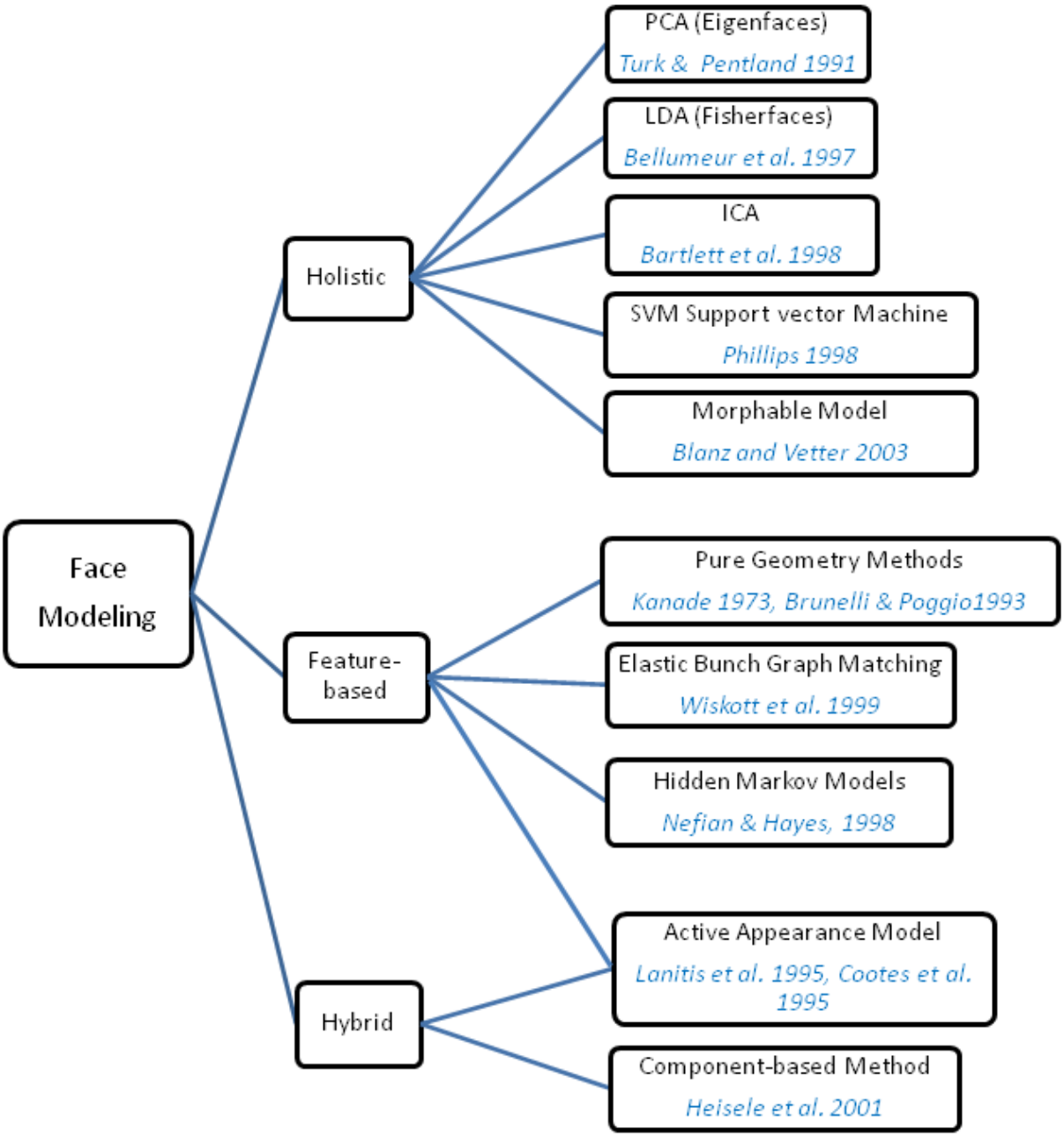

Figure 2: Feature extraction strategies. *Holistic:* use the entire face region in the recognition process; *Feature-based:* local features are extracted and fed into a classifier; *Hybrid:* use both local features and the holistic extraction.

Nonlinear subspace methods have also been proposed e.g. Kernel PCA [47], Isomap [48] and Locally Linear Embedding [49] that claim to model more accurately the nonlinear facial motions. However, intensive computing effort and a large amount of training data are usually needed to achieved good results, making these methods unsuitable for real-time applications such as biometrics [50].

2.2 Feature-based Approach

Feature-based approaches consist of locating specific facial features such as the eyes, the nose, the mouth, whose characteristics are then fed into a classifier for recognition [38]. The choice of the facial features depends on the types of motions, e.g. nose movements are important in an emotion of disgust, while eyebrow movements are predominant in emotions such as surprise, fear, and anger [18].

Some very popular works that adopt this approach include the work of Kanade [51] and Brunelli et al. [52] that measure facial geometric features for recognition. Also very successful is the Elastic Bunch Graph Matching [53] where faces are represented as graphs, with nodes positioned at fiducial points (e.g. lip corners, nose tip) and described by a set of Gabor wavelet coefficients. The geometry of the face is encoded by the edges that are labeled with 2D distance values. To identify a new face, an initial face graph is positioned on the face image, and the fiducial points of the query image are located by solving an optimisation problem.

2.3 Hybrid Methods

It has been suggested that, in a feature-based approach, a few dozens of fiducial points are insufficient to capture all the facial subtleties such as wrinkles and muscle folds. This observation has led to the development of hybrid approaches that use both holistic and local features, for instance, the concept of Eigenfaces can

be extended to features such as Eigeneyes, Eigenmouth [54] and used concurrently to enhance recognition. Methods such as the ASM and the AAM can also be extended to the hybrid approach and used in combination with feature-based matching. However, experiments have shown so far very small improvements compared to the holistic approach or the feature-based approach alone [55].

References

[1] P. Ekman and W. Friesen, "The Facial Action Coding System: A Technique for the Measurement of Facial Action," Consulting Psychologists (1978).

[2] K. Schmidt and J. Cohn, "Dynamics of facial expression: Normative characteristics and individual differences," IEEE International Conference on Multimedia and Expo pp. 728–731 (2001).

[3] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman, "Individual Differences in Facial Expression: Stability over Time, Relation to Self-Reported Emotion, and Ability to Inform Person Identification," in Proc of the International Conference on Multimodal User Interfaces **vol. 116,** pp. 491–498 (2002).

[4] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition **vol. 2,** pp. 568–573 (2005).

[5] J. Luettin, "Visual speech and speaker recognition," Ph.D. thesis, University of Sheffield, UK (1997).

[6] M. I. Faraj and J. Bigun, "Audio-visual person authentication using lip-motion from orientation maps," Pattern Recognition Letters **vol. 28(11),** pp. 1368–13824 (2007).

[7] M. Leszczynski and W. Skarbek, "Viseme recognition - a comparative study," in Proc. of the IEEE Conference on Advanced Video and Signal Based Surveillance pp. 287–292 (2005).

[8] H. L. Somers, "Similarity metrics for aligning children's articulation data," in Proc. of the 36th Annual Meeting of the Association for Computational Linguistics **vol. 2,** pp. 1227–1232 (1998).

[9] D. A. Fidaleo and M. Trivedi, "Manifold analysis of facial gestures for face recognition," ACM SIGMM Multimedia Biometrics Methods and Application Workshop (2003).

[10] G. Shakhnarovich and B. Moghaddam, "Face Recognition in Subspaces," Handbook of Face Recognition, Eds. Stan Z. Li and Anil K. Jain, Springer-Verlag (2004).

[11] T. Collins, "Facial Dynamics for Identity Recognition. PhD Thesis Proposal, supervisor Pr. B. Fisher," Institute for Perception, Action and Behaviour. School of Informatics. University of Edinburgh (2006).

[12] T. Lukins and R. Fisher, "Qualitative characterization of deforming surfaces," in Proc. of the Third International Symposium on 3D Data Processing, Visualization and Transmission (2006).

[13] S. Pamudurthy, E. Guan, K. Mueller, and M. Rafailovich, "Dynamic approach for face recognition using digital image skin correlation," Audio and Video-Based Biometric Person Authentication **vol. 3546,** pp. 1010–1018 (2005).

[14] E. Y. Zhang, S. J. Kundu, D. B. Goldgof, S. Sarkar, and L. V. Tsap, "Elastic Face, An Anatomy-Based Biometrics Beyond Visible Cue," in Proc. of International Conference on Pattern Recognition **vol. 2,** pp. 19–22 (2004).

[15] K. Waters, "A muscle model for animating three-dimensional facial expressions," Computer Graphics **vol. 21(4),** pp. 17–24 (1987).

[16] I. A. Essa and A. Pentland, "A vision system for observing and extracting facial action parameters," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition pp. 76–83 (1994).

[17] D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," in IEEE Transactions on Pattern Analysis and Machine Intelligence **vol. 15(6),** pp. 569–579 (1993).

[18] C. Pelachaud, N. I. Badler, and M. Viaud, "Final Report to NSF of the Standards for Facial Animation Workshop," Technical Report (1994).

[19] K. Waters and J. Frisbie, "A coordinated muscle model for speech animation," in Proc. of Graph. Interface pp. 163–170 (1995).

[20] L. Reveret and I. Essa, "Visual coding and tracking of speech related facial motion," In Proc. of IEEE CVPR Intl. Workshop on Cues in Communication pp. 163–170 (2001).

[21] S. Basu, N. Oliver, and A. Pentland, "3D modeling and tracking of human lip motions," IEEE Computer Society pp. 337–343 (1998).

[22] E. Sifakis, I. Neverov, and R. Fedkiw, "Automatic determination of facial muscle activations from sparse motion capture marker data," ACM Trans. on Graphics **vol. 24(3),** pp. 417–425 (2005).

[23] A. Yilmaz, K. Shafique, and M. Shah, "Estimation of rigid and non-rigid facial motion using anatomical face model," in Proc. of the 16th International Conference on Pattern Recognition **vol. 1,** pp. 377–380 (2002).

[24] K. Kahler, J. Haber, H. Yamauchi, and H. P. Seidel, "Generating animated head models with anatomical structure," in Proc. of the ACM SIGGRAPH Symposium on Computer Animation pp. 55–64 (2002).

[25] Y. Zhang, E. C. Prakash, and E. Sung, "Efficient modeling of an anatomybased face and fast 3D facial expression synthesis," in Computer Graphics Forum **vol. 22(2),** pp. 159–170 (2003).

[26] Y. Yacoob and L. S. Davis, "Computing spatio-temporal representations of human faces," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition pp. 70–75 (1994).

[27] K. Mase and A. Pentland, "Automatic lipreading by optical flow analysis," Systems and Computers in Japan **vol. 22(6),** pp. 67–76 (1991).

[28] D. DeCarlo and D. N. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," in International Journal of Computer Vision **vol. 38(2),** pp. 99–127 (2000).

[29] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in Proc. of the First International Conference on Computer Vision **vol. 1(4),** pp. 321–331 (1987).

[30] C. Bregler and Y. Konig, "Eigenlips for robust speech recognition," in Proc. of the Intl. Conf. on Acoustics Speech and Signal Processing (IEEE-ICASSP) **vol. 2,** pp. 669–672 (1994).

[31] M. Rydfalk, "Candide: A parameterized face," PhD Thesis, Linkoping University (1978).

[32] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," Computer Vision, Graphics and Image Understanding **vol. 1(61),** pp. 38–59 (1995).

[33] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," In Proc. European Conf. on Computer Vision **vol. 2,** pp. 484–498 (1998).

[34] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," IEEE Transactions on Pattern Analysis and Machine Intelligence **vol. 25(9),** pp. 1063–1074 (2003).

[35] M. Turk and A. Pentland, "Eigenfaces for recognition," Cognitive Neurosciences **vol. 3(1),** pp. 71–86 (1991).

[36] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence **vol. 19(7),** pp. 711–720 (1997).

[37] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Transactions on Neural Networks **vol. 13(6),** pp. 1450–1464 (2002).

[38] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," Pattern Recognition **vol. 35(4),** pp. 399–458 (2003).

[39] K. Delac, M. Grgic, and P. Liatsis, "Appearance-based statistical methods for face recognition," in Proc.of the 47th International Symposium ELMAR (2005).

[40] X. Lu, "3D Face Recognition across Pose and Expression," PhD Thesis. Michigan State University (2006).

[41] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face Recognition by Humans: 20 Results all Computer Vision Researchers Should Know About," Proceedings of the IEEE **vol. 94,** pp. 19481962 (2006).

[42] 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).

[43] P. J. Phillips, "Support vector machines applied to face recognition," in Proc of the Conf on Advances in Neural Information Processing Systems **vol. II,** pp. 803–809 (1999).

[44] J. Beveridge, S. K., B. Draper, and G. Givens, "A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition pp. 535–542 (2001).

[45] P. Navarrete and J. R. del Solar, "Analysis and Comparison of Eigenspace-Based Face Recognition Approaches," International Journal of Pattern Recognition and Artificial Intelligence **vol. 16,** pp. 817–830 (2002).

[46] K. Baek, B. Draper, J. Beveridge, and K. She, "PCA vs. ICA: A Comparison on the FERET Data Set," Proc. of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing pp. 824–827 (2002).

[47] B. Scholkopf, A. Smola, and K. Mueller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," Neural Computation **vol. 10(5),** pp. 1299–1319 (1998).

[48] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A Global geometric Framework for Nonlinear Dimensionality Reduction," Science **vol. 290,** pp. 2319–2323 (2000).

[49] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science **vol. 290(5500),** pp. 2323–2326 (2000).

[50] S. Raudys and A. Jain, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," IEEE Transactions on Pattern Analysis and Machine Intelligence **vol. 13(3),** pp. 252–264 (1991).

[51] T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces," PhD Thesis, University of Kyoto (1973).

[52] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," IEEE Transactions on Pattern Analysis and Machine Intelligence **vol. 15(10),** pp. 1042–1052 (1993).

[53] L. Wiskott, J. M. Fellous, N. Krueuger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," Chapter 11 in Intelligent Biometric Techniques in Fingerprint and Face Recognition pp. 355–396 (1995).

[54] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition pp. 84–91 (1994).

[55] W. Zhao and R. Chellappa, "Face Processing: Advanced Modeling and Methods," in Academic Press, Sarnoff Corporation, Princeton, NJ, USA (2006).