Ongoing Challenges in Face Recognition

Peter N. Belhumeur
Department of Computer Science
Columbia University
New York, New York 10027
belhumeur@cs.columbia.edu

Figure 1: The same individual imaged with the same camera and seen with nearly the same facial expression and pose may appear dramatically different with changes in the lighting conditions. The two leftmost images were taken indoors and the two rightmost were taken outdoors. All four images were taken with a Canon EOS 1D digital camera. Before each acquisition the subject was asked to make a neutral facial expression and to look directly into the lens.

It has been observed that "the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity" [46]. As is evident in Figures 1 and 2, the same person, with the same facial expression, can appear strikingly different when light source direction and viewpoint vary. These variations are made even greater by additional factors such as facial expression, perspiration, hair styles, cosmetics, and even changes due to aging.

The problem of face recognition can be cast as a standard pattern classification or machine learning problem: Given a set of face images labeled with the person's identity (the gallery set) and an unlabeled set of face images from the same group of people (the probe set), we seek to identify each person in the probe images. This problem is attacked in three steps. In the first step, the face is located in the image; this process, known as face detection, is in many respects as challenging a problem as face recognition, see [68b, 70b] for more detail. In the second step, a collection of descriptive measurements known as a feature vector is extracted from each image. In the third step, a classifier is trained to assign to each feature vector a label with a person's identity. (Note that these classifiers are simply mathematical functions which, given a feature vector, return an index corresponding to a subject's identity.)

Over the last few years, numerous feature extraction and pattern classification methods have proposed for face recognition, see surveys [59, 7, 17, 50, 77]. For decades, geometric feature-based methods [21, 33, 35, 34, 27, 26, 59, 6, 69, 42] have used properties and relations (e.g., distances and angles) between facial features such as eyes, mouth, nose, and chin to perform recognition. Despite their economical representation and their insensitivity to small variations in illumination and viewpoint, feature-based methods are quite sensitive to the feature extraction and measurement process. It has been argued that existing techniques for the extraction and measurement of facial features are not reliable enough [12]. It has also been claimed that methods for face recognition based on finding local image features and inferring identity by the geometric relations of these features are often ineffective [6].

Methods have been introduced over the last decade that use low-dimensional representations of images of objects or faces to perform recognition. See for example [36, 66, 23, 51, 55, 47, 45, 25]. These methods, often termed appearance-based methods, differ from feature-based techniques in that their low-dimensional representation is, in a least-squares sense, faithful to the original image. Techniques such as SLAM [47] and Eigenfaces [66] have demonstrated the power of appearance-based methods both in ease of implementation and in accuracy. Here the feature vector used for classification is a linear projection of the face image into a lower-dimensional linear subspace. In extreme cases, the feature vector is chosen as the entire image, with each element in the feature vector taken from a pixel in the image.

Despite their success, many of the appearance-based methods suffer from an important drawback: recognition of a face under a particular lighting condition, pose, and expression can be performed reliably ***provided the face has been previously seen under similar circumstances***. It is precisely this variation in appearance between images of the same person that confounds appearance-based methods. To demonstrate just how severe this variability can be, an array of images is included showing variability in the Cartesian product of pose x lighting for a single individual, see again Figure 2.

If the gallery set were to contain a very large number of images of each subject taken from many poses, under many lighting conditions, and with many facial expressions, then one might expect that even the simplest appearance-based classifier might perform well. In the deployment of face recognition systems, however, there are usually only a few gallery images per person from which the classifier must learn to discriminate between individuals.

In an effort to overcome this obstacle, there has been a recent surge in work on 3-D face recognition. The idea here is build face recognition systems that use a handful of images acquired at enrollment time to estimate models of the 3-D shape of each face. These 3-D models can then be used to synthetically render images of each face under novel pose and lighting conditions – effectively increasing the gallery set for each face. Alternatively, these models can be used in an iterative fitting process in which the model for each face is rotated, aligned, and synthetically illuminated to best match the probe image. Conversely, the models can be used to warp a probe image of a face back to a canonical frontal viewpoint and lighting condition. In both cases the identity is chosen as that corresponding to the model with the best fit.

The 3-D models of the face shape can be estimated by a variety of methods. In the simplest of methods the face shape is assumed to be a generic average of a large collection of sample face shapes acquired from laser range scans. In [18b, 19], the face shape is estimated from the shading changes in multiple enrollment images of the same face seen under varying lighting conditions. In [35b], the shape is estimated using binocular stereopsis on two enrollment images taken from slightly different view points. In [48b], the shape is estimated from deformations in grid pattern of IR light projected onto the face. In [5c], the 3-D face shape is inferred from the shading in a single image using a parametric model of face shape. In much of the 3-D face recognition work, a bootstrap set of prior training data of face shape and reflectance, taken from individuals not in either the gallery or probe sets, is used to better the shape and reflectance estimation process.

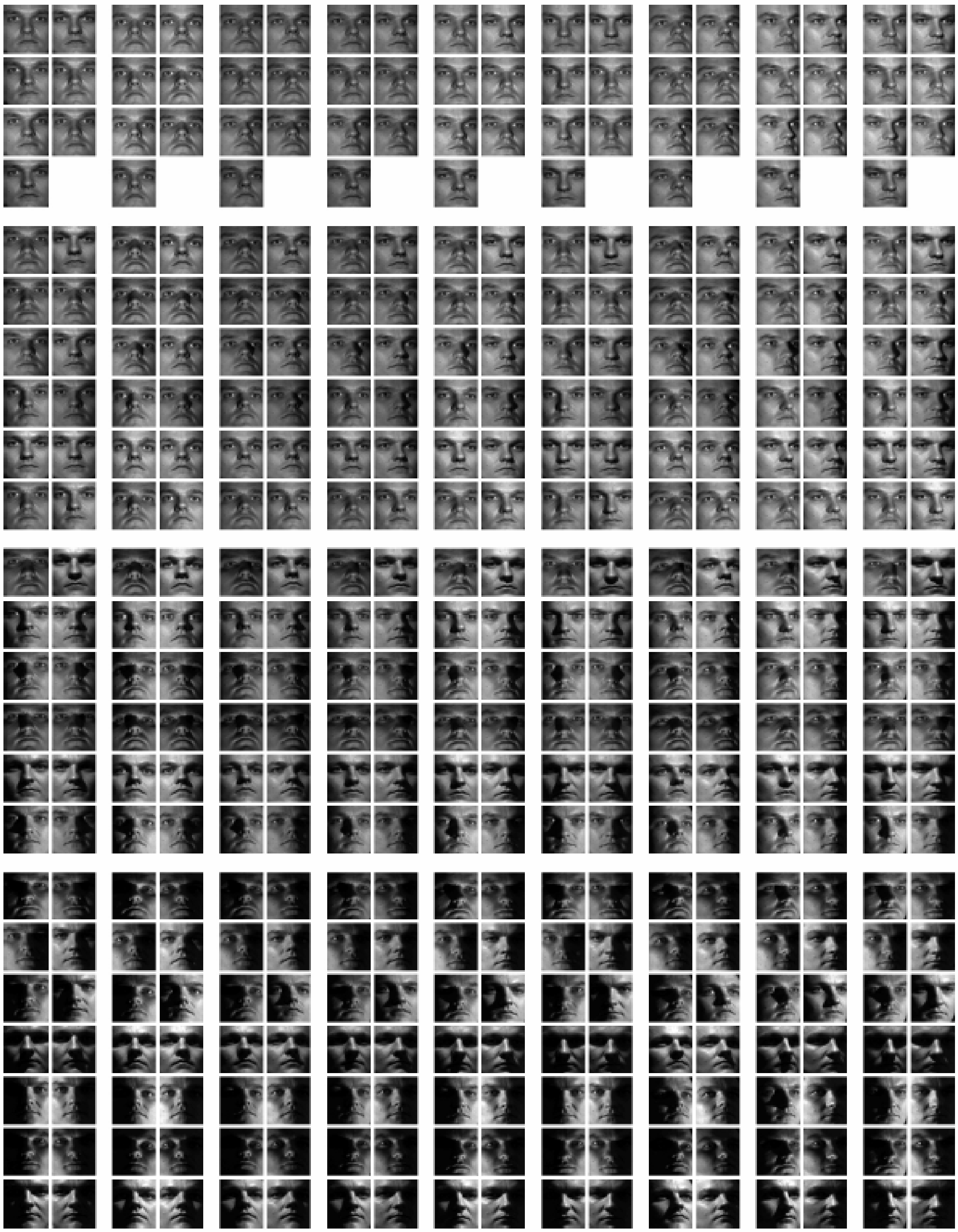

Figure 2: Images of a single individual from the Yale Face Database B showing the variability due to illumination and pose. The images have been divided into four subsets (1 through 4 from top to bottom) according to the angle the light source direction makes with the camera axis. Every pair of columns shows the images of a particular pose (1 through 9 from left to right).

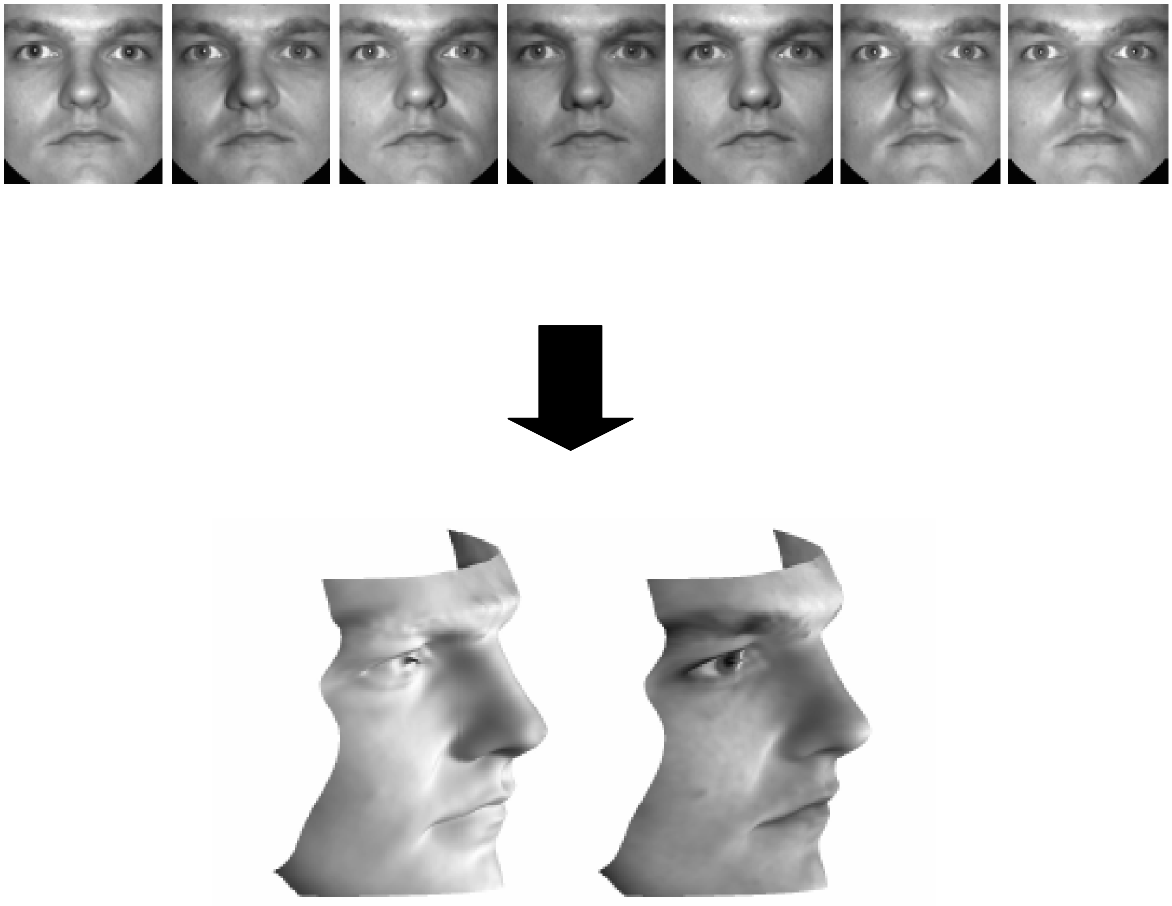

Figure 3: The top row shows seven gallery images of the face used to compute the shape and reflectance of the face as shown in the bottom row. The method used to do this is a variant of what is known as photometric stereopsis, see [18b] for details.

The above list is just a small sampling of the work that is going on in this area, much of it too new to appear in surveys. To see the potential of these approaches, we show results from [18b]. Here seven images under variable lighting are used to estimate the face shape and reflectance and this is in turn used to synthesize images of the face shown in Figure 4 under the same conditions shown in Figure 2. Note that much of the appearance variation in pose and lighting can be inferred from as few as nine gallery images.

Yet, while recent advances in 3-D face recognition have gone a long way toward addressing complications due to pose and lighting, much remains to be done. Natural outdoor lighting has proven difficult, not simply because of the strong shadows cast by a light source such as the sun, but also because subjects tend to distort their faces when illuminated by a strong source; compare again the indoor and outdoor expressions of the subject in Figure 1. Furthermore, there has been very little work to address complications arising from voluntary changes in facial expression, the use of eyewear, and the more subtle effects of aging. The hope, of course, is that many of these effects can be modeled in much the same manner as face shape and reflectance and that recognition performance will continue to improve over the coming decade.

Figure 4: Synthesized images of the same individual and under the same illumination and viewpoints as in Figure 2. As before, the synthesized images have been divided into four subsets (1 through 4 from top to bottom) according to the angle the light source direction makes with the camera axis. Every pair of columns shows the images from a particular pose (1 through 9 from left to right). Note that all the images were generated synthetically from seven gallery images with frontal pose.

[1] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions. *Int. J. Computer Vision*, 28(3):245–260, July 1998.

[2] P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 1040–1046, 1997.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):711–720, 1997. Special Issue on Face Recognition.

[4] D. Beymer. Face recognition under varying pose. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 756–761, 1994.

[5] D. Beymer and T. Poggio. Face recognition from one example view. In *Int. Conf. on Computer Vision*, pages 500–507, 1995.

[5b] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and interactive Techniques* International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, pages 187-194.

[5c] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 9, pages 1063—1074, 2003.

[6] R. Brunelli and T. Poggio. Face recognition: Features vs templates. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 15(10):1042–1053, 1993.

[7] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

[8] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 254–261, 2000.

[9] Q. Chen, H. Wu, and M. Yachida. Face detection by fuzzy pattern matching. In *Int. Conf. on Computer Vision*, pages 591–596, 1995.

[10] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498, 1998.

[11] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[12] I. Cox, J. Ghosn, and P. Yianilos. Feature-based face recognition using mixture distance. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 209–216, 1996.

[13] I. Craw, D. Tock, and A. Bennet. Finding face features. pages 92–96, 1992.

[14] G. Edwards, T. Cootes, and C. Taylor. Advances in active appearance models. In *Int. Conf. on Computer Vision*, pages 137–142, 1999.

[15] R. Epstein, P. Hallinan, and A. Yuille. 5+/-2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *Physics Based Modeling Workshop in Computer Vision*, Session 4, 1995.

[16] R. T. Frankot and R. Chellapa. A method for enforcing integrabilty in shape from shading algorithms. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 10(4):439–451, 1988.

[17] T. Fromherz. Face recognition: A summary of 1995-1997. International Computer Science Institute ICSI TR- 98-027, University of California, Berkeley, 1998.

[18] A. Georghiades, P. Belhumeur, and D. Kriegman. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting. In *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, pages 47–54, Fort Collins, Colorado, 1999.

[18b] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 6 (Jun. 2001), pages 643-660.

[19] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 277–284, 2000.

[20] A. Georghiades, D. Kriegman, and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 52–59, 1998.

[21] A. Goldstein, L. Harmon, and A. Lesk. Identification of human faces. *Proceedings of the IEEE*, 59(5):748–760, May 1971.

[22] V. Govindaraju. Locating human faces in photographs. *Int. Journal of Computer Vision*, 19(2):129–146, August 1996.

[23] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 995–

999, 1994.

[24] P. Hallinan. *A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions*. PhD thesis, Harvard University, 1995.

[25] P. Hallinan, G. Gordon, A. Yuille, and D. Mumford. *Two- and Three-Dimensional Patterns of the Face*. A.K.Peters, 1999.

[26] L. Harmon, M. Kaun, R. Lasch, and P. Ramig. Machine identification of human faces. *Pattern Recognition*, 13(2):97–110, 1981.

[27] L. Harmon, S. Kuo, P. Ramig, and U. Raudkivi. Identification of human face profiles by computer. *Pattern Recognition*, 10:301–312, 1978.

[28] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *J. Opt. Soc. Am. A*, 11(11):3079–3089, Nov. 1994.

[29] B. Horn. *Computer Vision*. MIT Press, Cambridge, Mass., 1986.

[30] F. Huang, Z. Zhou, H. Zhang, and T. Chen. Pose invariant face recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 245–250, 2000.

[31] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and characterizing intensity images. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1997.

[32] P. Juell and R. Marsh. A hierarchical neural-network for human face detection. *Pattern Recognition*, 29(5):781–787, May 1996.

[33] T. Kanade. *Picture Processing by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973.

[34] T. Kanade. *Computer Recognition of Human Faces*. Birkhauser Verlag, Stuttgart, Germany, 1977.

[35] G. Kaufman and K. Breeding. The automatic recognition of human faces from profile silhouettes. *IEEE Trans. on Systems, Man and Cybernetics*, 6:113–121, February 1976.

[35b] J. Kukula. Effects of light direction on the performance of Geometrix FaceVision® 3D face recogntion system. Biometrics Consortium Conference, 2004.

[36] L. Sirovitch and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Optical Soc. of America A*, 2:519–524, 1987.

[37] J. Lambert. *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae*. Eberhard Klett, 1760.

[38] A. Lanitis, C. Taylor, and T. Cootes. A unified approach to coding and interpreting face images. In *Int. Conf. on Computer Vision*, pages 368–373, 1995.

[39] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):743–756, July 1997.

[40] C. Lee, J. Kim, and K. Park. Automatic human face location in a complex background using motion and color information. *Pattern Recognition*, 29(11):1877–1889, November 1996.

[41] T. Leung, M. Burl, and P. Perona. Finding faces in cluttered scenes using labeled random graph matching. In *Int. Conf. on Computer Vision*, pages 637–644, 1995.

[42] S. Li and J. Lu. Face recognition using nearest feature line. *IEEE Trans. on Neural Networks*, 10(2):439–443, March 1999.

[43] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–305, 2000.

[44] Moghaddam and Pentland. Probabilistic visual learning for object detection. In *Int. Conf. on Computer Vision*, pages 786–793, 1995.

[45] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):696–710, 1997.

[46] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In *European Conf. on Computer Vision*, pages 286–296, 1994.

[47] H. Murase and S. Nayar. Visual learning and recognition of 3-D objects from appearence. *Int. J. Computer Vision*, 14(5–24), 1995.

[48] S. Nayar and H. Murase. Dimensionality of illumination manifolds in appearance matching. In *Int. Workshop on Object Representations for Computer Vision*, page 165, 1996.

[48b] F. Ohlhorst. Biometrics security solutions are here from the future. http://www.crn.com/showArticle.jhtml?articleID=161601680&flatPage=true. April 26, 2005.

[49] A. O'Toole, T. Vetter, N. Toje, and H. Bulthoff. Sex classification is better with three-dimensional head structure than with texture. *Perception*, 26:75–84, 1997.

[50] A. Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 22:107–119, 2000.

[51] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 84–91, 1994.

[52] P. Phillips, H. Moon, P. Rauss, and S. Risvi. The FERET evaluation methodology for face-recognition algorithms. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 137–143, 1997.

[53] P. Phillips, H. Moon, P. Rauss, and S. Risvi. The FERET September 1996 database evaluation procedure. In *Audio and Video-Based Biometric Person Authentication*, 1997.

[54] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for facerecognition algorithms. *Image and Visual Computing*, 16(5), 1998.

[55] T. Poggio and K. Sung. Example-based learning for view-based human face detection. In *Proc. Image Understanding Workshop*, pages II:843–850, 1994.

[56] T. Riklin-Raviv and A. Shashua. The quotient image: Class based recognition and synthesis under varying illumination conditions. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages II:566–571, 1999.

[57] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 20(1):23–38, January 1998.

[58] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 38–44, 1998.

[59] A. Samil and P. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25:65–77, 1992.

[60] A. Samil and P. Iyengar. Human face detection using silhouettes. *Pattern Recognition and Artificial Intelligence*, 9:845–867, 1995.

[61] A. Shashua. *Geometry and Photometry in 3D Visual Recognition*. PhD thesis, MIT, 1992.

[62] A. Shashua. On photometric issues to feature-based object recognition. *Int. J. Computer Vision*, 21:99–122, 1997.

[63] H. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 17(9):854–867, September 1995.

[64] W. Silver. *Determining Shape and Reflectance Using Multiple Images*. PhD thesis, MIT, Cambridge, MA, 1980.

[65] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. J. Computer Vision*, 9(2):137–154, 1992.

[66] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1):71–96, 1991.

[67] T. Vetter. Synthesis of novel views from a single face image. *Int. Journal of Computer Vision*, 28(2):103–116, June 1998.

[68] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):733–742, July 1997.

[68b] P. Viola, M. J. Jones. Robust Real-Time Face Detection. *Int. J. Comput. Vision* 57, 2, pages 137—154. May 2004.

[69] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):775–779, July 1997.

[70] R. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 17:117–140, 1981.

[70b] M. Yang, D. J. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 1, pages 34—58. January 2002.

[71] M. Yang, N. Ahuja, and D. Kriegman. Mixture of linear subspaces for face detection. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 196–201, 2000.

[72] K. Yow and R. Cipolla. Feature-based human face detection. *Image Visual Computing*, 15(9):713–735, September 1997.

[73] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 207–218, 1998.

[74] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 158–164, 1997.

[75] W. Zhao and R. Chellapa. SFS based view synthesis for robust face recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 285–292, 2000.

[76] W. Zhao, R. Chellappa, and P. Phillips. Subspace linear discriminant analysis for face recognition. Center for Automation Research CAR-TR-914, University of Maryland, College Park, 1999.

[77] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.* 35, 4, pages 399—458, 2003.