OVERVIEW OF FACE RECOGNITION TECHNIQUES

Ioan Buciu
Department of Electronics,
University of Oradea, Faculty of Electrical Engineering and Information Technology,
Universitatii, 1, 410087, Oradea, Romania, E-Mail: ibuciu@uoradea.ro

***Abstract* – Humans are able to rapidly and accurately recognize familiar faces and identities under widely varying and difficult viewing conditions, such as illuminations changing, occlusion, scaling or rotation. Motivated by its importance in human-to-human communication and leading to various applications, from biometrics to human – computer interaction, the face recognition task is a major issue in the Computer Vision field and more. This paper presents a brief survey on state-of-the art methods used to cope with this challenging task.**

***Keywords:* face recognition techniques, technologies, applications, open problems.**

I. INTRODUCTION

Among other applications, face recognition is one of the primary biometric tasks, becoming more and more important as advances in technologies such as Internet, digital cameras require increased security key features. Face recognition can operate either on still images or image sequences. Also, it can manage either or both tasks: face identification (or recognition) and face verification (or authentication). Face recognition is extreme challenging task as the recognition performance significantly deteriorates with changes in lighting, pose, scale or occlusion. The paper is structured as follows. Section II presents ones of the most representatives techniques developed for subspace analysis. It also deals with the face recognition solutions under varying pose and illumination conditions. The Section ends up with the face recognition methods for image sequences. Section III describes several face recognition applications. Finally, major limitations of the existing face recognition systems are mentioned in Section IV.

II. FACE RECOGNITION APPROACHES

A. Face Recognition in Subspaces

Subspace analysis methods rely on the hypothesis that the face images resides in a lower subspace of the input image space. Thus, many information captured in the input space is redundant from the face pattern point of view. The features contained in such subspace provide richer information for face recognition. Probably the most used subspace approach is given by the Principal Component Analysis (PCA) that extracts features named eigenfaces [1] representing a face image. PCA represents faces by their projection onto a set of orthogonal axes (also known as principal components, eigenvectors, eigenfaces, or basis images) pointing into the directions of maximal covariance in the facial image data. By defining the covariance matrix with

$$\mathbf{C}_x = E\left\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T\right\}$$

where $\boldsymbol{\mu}_x$ denotes the mean image. PCA solution is found by solving the equations system

$$\mathbf{C}_x\mathbf{Z}_{PCA} = \boldsymbol{\lambda}\mathbf{Z}_{PCA}$$

with $\boldsymbol{\lambda}$ as eigenvectors. The basis images corresponding to PCA are typically ordered according to the decreasing amount of variance they represent, i.e., the respective eigenvalues. Here ZPCA comprises the eigenfaces. The PCA techniques essentially select a subspace that retains most of that variation, and consequently the similarity in the face space is not necessarily determined by the identity. To overcome this, Belhumeur et al. [2] developed so called "Fisherfaces", an application of Fisher's linear discriminant (FLD). This technique projects the images onto a subspace where the classes are maximally separated by maximizing the between-classes scatter matrix and minimizing the within-class scatter matrix at the same time. If we denote the set of all N = | x | data divided into *c* classes with $X \equiv \{x_1, x_2, ..., x_c\}$, then the inter-class scatter matrix $\mathbf{S}_w$ is defined as

$$\mathbf{S}_w = \sum_{i=1}^{c}\sum_{x_k \in x_i}(x_k-\mu_i)(x_k-\mu_i)^T$$

while the between-class scatter matrix Sb is defined as

$$\mathbf{S}_b = \sum_{i=1}^{c}|\,x_i\,|\,(\mu_i-\mu)(\mu_i-\mu)^T$$

where μ_i is the mean image of class xi and μ is the mean of all data. Here, ZFLD satisfies

$$\mathbf{Z}_{FLD} = \arg\max_{Z}\frac{\left|\mathbf{Z}^T\mathbf{S}_B\mathbf{Z}\right|}{\left|\mathbf{Z}^T\mathbf{S}_w\mathbf{Z}\right|}$$

The solution for finding ZFLD is to solve the generalized eigeinvalues problems:

$$\mathbf{S}_B \mathbf{Z}_{LDA} = \boldsymbol{\lambda} \mathbf{S}_w \mathbf{Z}_{LDA}$$

While PCA minimizes the sample covariance (second-order dependence) of the data, independent component analysis (ICA) [3] minimizes higher-order dependencies as well, and the components found by ICA are designed to be non-Gaussian. Bartlett et al. [3] used two ICA configurations to represent faces for recognition. PCA was carried out prior to ICA for dimensionality reduction. An intermediate step for "whitening" the data has been introduced between PCA and ICA processing. The data were then decomposed into basis images and de-composition coefficients. Their second ICA configuration (ICA2) yields holistic basis images very similar to those produced by PCA. In that case, ICA is applied to the projection matrix containing the principal components. Under this architecture, the linear decomposition coefficients are as independent as possible. A nonlinear variant of subspace analysis was prposed by Vasilescu and Terzopoulos [4] called "Tensorfaces". Finally, kernel – based approaches were poposed, inclusing kernel PCA [5], kernel LDA [6]. In experiments on two data sets that contained images from 40 and 11 subjects, respectively, with varying pose, scale, and illumination, this algorithm showed performance clearly superior to that of ICA, PCA, and KPCA and somewhat better than that of the standard Fisherfaces.

B. Face Recognition Across Pose and Illumination

The most recent evaluation of commercial face recognition systems shows satisfactory performance for face verification of the best systems to be on par with fingerprint recognizers for frontal, uniformly illuminated faces. Recognizing faces across changes in pose and illumination has proved to be a much more difficult problem. Although most research has so far focused on frontal face recognition, there is a sizable body of work on pose invariant face recognition and illumination invariant face recognition. However, face recognition across pose and illumination has received little attention. One of the earliest appearance-based multi-view approaches was the one developed by Beymer [7]. After a pose estimation step, the algorithm geometrically aligns the probe images to candidate poses of the gallery subjects using the automatically determined locations of three feature points. This alignment is then refined using the optical flow strategy. Recognition is performed by computing normalized correlation scores. Good recognition results are reported on a database of 62 subjects imaged in a number of poses ranging from -30^0 to $+30^0$ (yaw) and from -20^0 to $+20^0$ (pitch). Pentland et al. [8] extended the eigenface approach to tackle multiple views. The authors compare the performance of a parametric eigenspace (computed using all views from all subjects) with view-based eigenspaces (separate eigenspaces for each view). In experiments involving 21 people recorded in nine evenly spaced views from minus 90^0 to $+90^0$, view-based eigenspaces outperformed the parametric eigenspace by a small margin. A number of 2D model-based algorithms have been proposed for face tracking through large pose changes. Separate active appearance models [9] were trained for profile, half-profile, and frontal views, with models for opposing views created by simple reflection. Using a heuristic for switching between models, the system was able to track faces through wide angle changes. It has been shown that linear models are able to deal with considerable pose variation so long as all the modeled features remained visible [10]. A different way of dealing with larger pose variations is then to introduce nonlinearities into the model. Romdhani et al. extended active shape models [11] and active appearance models [12] using a kernel PCA to model shape and texture nonlinearities across views. In both cases models were successfully fit to face images across a full 180^0 rotation. However, no face recognition experiments were reported.

C. Face Recognition From Image Sequences

Recognizing faces in image sequences (video) is a more difficult task than the one corresponding to recognizing face in still images as it involves simultaneously tracking and recognition. Typically, a video-based face recognition system operates as follows. The face is first detected and then tracked over time. Only when a frame satisfying certain criteria (size, pose) is acquired is recognition performed using still-to-still recognition technique. For this, the face part is cropped from the frame and transformed or registered using appropriate transformations. This tracking-then-recognition approach attempts to resolve uncertainties in tracking and recognition sequentially and separately and requires a criterion for selecting good frames and estimation of parameters for registration. Probabilistic video analysis has recently gained significant attention in the computer vision community since the work of Isard and Blake [13]. The authors introduced a time series state space model parameterized by a tracking motion vector (e.g., affine transformation parameters). The CONDENSATION algorithm was developed to provide a numerical approximation to the posterior distribution of the motion vector at time t given the observations up to t. The CONDENSATION algorithm, also known as the particle filter, was originally proposed in the signal processing literature [14] and has been used to solve many other vision tasks [15, 16], including human face recognition [17].

III. FACE RECOGNITION APPLICATIONS

This Section briefly reviews major face recognition applications.

1. *Face identification* was widely used for identifying driver licenses, in immigration programs, passports, or welfare registration.
2. *Access control* deals with border – crossing, vehicle access, ATM, computer access, computer network access, online transaction access, online

database access. For example, a commercial access control named FaceGate [18] requires the one wishing to get into a building to enter his entry code or a card and face a camera on the door entry system. By applying a mathematical model to an image of a face, FaceGate generates a unique biometric "key." Whenever one wishes to access a building, FaceGate verifies the person's entry code or card, then compares his face with its stored "key." It registers him as being authorized and allows him to enter the building. Access is denied to anyone whose face does not match.

3. *Security* refers to the terrorist alert issue in airports, secure boarding systems, file encryption, intranet and internet security, or medical records. Many airports have been adopted face recognition technology for improving security. In contrast to face identification issue, face-recognition based security is a harder task due to many factors, such as different lighting conditions which can not always be controlled, or the large number of faces that have to be examined in public places and to discriminate the interested one from the crowd. For instance, Fresno Yosemite International (FYI) airport in California deployed Viisage's face recognition technology for airport security purposes. The system is designed to alert FYI's airport public safety officers whenever an individual matching the appearance of a known terrorist suspect enters the airport's security checkpoint. Anyone recognized by the system would undergo further investigative processes by public safety officers. Unfortunately, no satisfactory results (if not quite disappointing) were obtained using automatic face recognition systems after several years of real-life testing.
4. *Surveillance* is another application area where face recognition plays a major part, including video surveillance, CCTV control or portal control. As in face recognition applied in public places, surveillance based on face recognition systems assesses even lower performance. Unconstrained lighting conditions and large variability in pose degree make this task extremely difficult for large scale. It is worth mentioning that, in 1998 Visionics FaceIt technology was deployed for the first time to enhance town center surveillance in Newham Borough of London, which has 300 cameras linked to the closed circuit TV (CCTV) control room. The city council claims that the technology has helped to achieve a 34% drop in crime since its installation. Similar systems are in place in Birmingham, England. In 1999 Visionics was awarded a contract from National Institute of Justice to develop smart CCTV technology [19].
5. *Smart cards* have an embedded microprocessor or memory chip that provides the processing power to serve many applications. Memory cards simply store data. A microprocessor card, on the other hand, can add, delete, and manipulate information in its memory on the card. A microprocessor card also has built-in security features. Contact-less smart cards contain a small antenna so the card reader detects the card from a distance. The Smart Card's portability and ability to be updated make it a technology well suited for securely connecting the virtual and physical worlds. The application of face recognition technology in smart cards, in essence, is a combination of the two. This can be seen from the following two examples. Smart cards store the mathematical characteristics of the faces during the enrollment stage. The characteristics are read out during the verification stage for comparison with the live capture of the person's face. If granted, the person can have his stored facial characteristics updated in the card's memory. To mention only one such application, the ZN-Face system [20] combines face recognition and smart card technology, is used for protecting secure areas at Berlin airports. Potential threats posed by criminals who often succeed in entering high security areas by means of a suitable disguise (e.g., pilot uniforms) are ruled out effectively. The individual's face characteristics are stored on a smart card; ZN-Face compares and verifies the card information with the face readings at each access station.
6. *Multimedia management* deals with face-based searching information, face-based video segmentation and summarization or event detection. Human faces are frequently seen in news, sports, films, home video, and other multimedia content. Indexing this multimedia content by face detection, face tracking, face recognition, and face change detection is important to generate segments of coherent video content for video browsing, skimming, and summarization. Together with speech recognition, natural language processing, and other image understanding techniques, face processing is a powerful tool for automatic indexing, retrieval, and access to the ever-growing digital multimedia content. One integrated multimedia management system is the "Infomedia" project at Carnegie Mellon University [21]. This project aims to create an information digital video library to enhance learning for people of all ages. Thousands of hours of video content is indexed and archived for search and retrieval by users via desktop computers through computer networks. Face databases. Content-based image retrieval tries to solve the difficulties faced by text-based image retrieval. Instead of being manually annotated by text-based keywords, images would be indexed by their own visual content, such as color and texture. Feature vector is the basis of content-based image retrieval, which captures image properties such as

color and texture. However, these general features have their own limitations. Recently, researchers have tried to combine it with other image analysis technologies, such as face detection and recognition, to improve the retrieval accuracy.

7. *Low enforcement* is closely related to suspect tracking and investigation, identifying cheats in casinos, criminal face retrieval and recognition.
8. *Human – computer interaction* refers to interactive gamming and proactive computing
9. *Other* applications include antique photo verification, very – low bit rate image and video transmission, etc.

IV. LIMITATIONS OF CURRENT FACE RECOGNITION SYSTEMS

Despite tremendous work performed to build a reliable face recognition system, the existing face recognition systems face several limitations. Face recognition technology is still not robust, especially in unconstrained environments, and recognition accuracy is not acceptable, especially for large-scale applications. Lighting changes, pose changes, and time differences between the probe image and the gallery image(s) further degrade the performance. These factors have been evaluated in FRVT 2002 using some of the best commercial systems [22]. For example, in a verification test with reasonably controlled indoor lighting, when the gallery consisted of 37,437 individuals with one image per person and the probe set consisted of 74,854 probes with two images per person, the best three systems, on average, achieved a verification rate of 90% at a false alarm rate of 1%, 80% at a false alarm rate of 0.1%, and 70% at a false alarm rate of 0.01%. This level of accuracy may be (or may not be) suitable for an access control system with a small database of hundreds of people but not for a security system at airports where the number of passengers is much larger. The test results in FRVT 2002 can partly explain why several systems installed at airports and other public places have not received positive feedback based on their poor performance. One example is that the crowd surveillance system tested by Tampa, Florida police reported 14 instances of a possible criminal match in a 4-day session, but they were all false alarms. The Tampa police department has abandoned the system.

REFERENCES

[1] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4(3):519–524, 1987.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fish-erfaces: Recognition Using Class Specific Linear Projection”, IEEE Trans. On Pattern Analysis and Machine Intelligence, 19(7), pp. 711-720, 1997.

[3] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” IEEE Trans. Neural Networks, 13(6), pp. 1450-1464, 2002.

[4] M. A. O. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: TensorFaces. In Proceedings of European Conference on Computer Vision, pages 447–460, Copenhagen, Denmark, May 2002.

[5] B. Scholkopf, A. Smola, and K.-R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10(5):1299–1319, 1998.

[6] M.-H. Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In Proc. of IEEE Internation Conf. on Face and Gesture Recognition, pages 215–220, Washington, DC, May 2002

[7] D. Beymer. Face recognition under varying pose. Technical Report 1461, MIT AI Laboratory, Cambridge, MA, 1993.

[8] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 84–91, 1994.

[9] T. Cootes, G. Wheeler, K. Walker, and C. Taylor. View-based active appearance models. Image and Vision Computing, 20:657–664, 2002.

[10] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):743–756, 1997.

[11] S. Romdhani, S. Gong, and A. Psarrou. Multi-view nonlinear active shape model using kernel PCA. In 10th British Machine Vision Conference, volume 2, pages 483–492, 1999.

[12] S. Romdhani, A. Psarrou, and S. Gong. On utilising template and feature-based correspondence in multi-view appearance models. In 6th European Conference on Computer Vision, volume 1, pages 799–813, 2000.

[13] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. European Conference on Computer Vision, pages 343–356, 1996.

[14] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings on Radar and Signal Processing, 140:107–113, 1993.

[15] G. Qian and R. Chellappa. Structure from motion using sequential Monte Carlo methods. IEEE Internation Conference on Computer Vision, pages 614–621, 2001.

[16] M.J. Black and A.D. Jepson. A probabilistic framework for matching temporal trajectories. European Conference on Computer Vision, pages 909–924, 1998.

[17] B. Li and R. Chellappa. A generic approach to simultaneous tracking and verification in video. IEEE Transactions on Image Processing, 11:530–544, 2002.

[18] FaceGate.http://www.premierelect.co.uk/faceaccess.html.

[19] FaceIt-Hist. http://www.identix.com/company/comp history.html.

[20] W. Konen and E. Schulze-Kruger. ZN-face: a system for access control using automated face recognition. In Proceedings of the International Workshop on Automatic Face and Gesture Recognition, pages 18–23, 1995.

[21] H. Wactlar, T.K.M. Smith, and S. Stevens. Intelligence access to digital video: informedia project. IEEE Computer, 29-5:46–52, 1996

[22] P. Phillips, P. Grother, R. Michaels, D. Blackburn, E.Tabassi, and M. Bone. Face recognition vendor test 2002: evalution report. http://www.frvt2002.org.