Face Recognition by Computers and Humans

Rama Chellappa
Department of Electrical and Computer Engineering and
Center for Automation Research, UMIACS
University of Maryland, College Park, MD., 20742.

Pawan Sinha
Department of Brain and Cognitive
Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

and

P. Jonathon Phillips
National Institute of Standards and Technology
Gaithersburg, MD, 20899.

1. INTRODUCTION

In most situations, identifying humans using faces is an effortless task for humans. Is this true for computers? This very question defines the field of automatic face recognition, one of the most active research areas in computer vision, pattern recognition, and perception. Over the past two decades, the problem of face recognition has attracted substantial attention from various disciplines and has witnessed an impressive growth in basic and applied research, product development and applications. Face recognition systems have been deployed at ports of entry at international airports in Australian and Portugal. In addition, studies on human perception of faces [1] have resulted in many interesting findings that can be used in the design of practical systems. Besides applications related to identification and verification such as access control, law enforcement, ID and licensing, surveillance, etc., face recognition is also useful in human-computer interaction, virtual reality, database retrieval, multimedia, computer entertainment, etc. More detailed discussions on face acquisition, processing, recognition and verification may be found in a survey paper [2], research monographs and recently published edited books [3, 4]. In this paper, we begin with a discussion of why automatic face recognition is hard, present a brief review of the past two decades of work in face recognition and then present a brief outline of future research trends.

A schematic of a general face recognition system is given in Figure 1. It consists of three major modules: face detection, feature extraction and face recognition. Some of the sub-functionalities in each of these modules are also listed. As in any pattern recognition problems, variations in patterns due to illumination, pose, expressions, etc are handled either in the feature extraction stage by making the features invariant or robust to these transformations or in the recognition stage, by designating rules that account for these transformations.

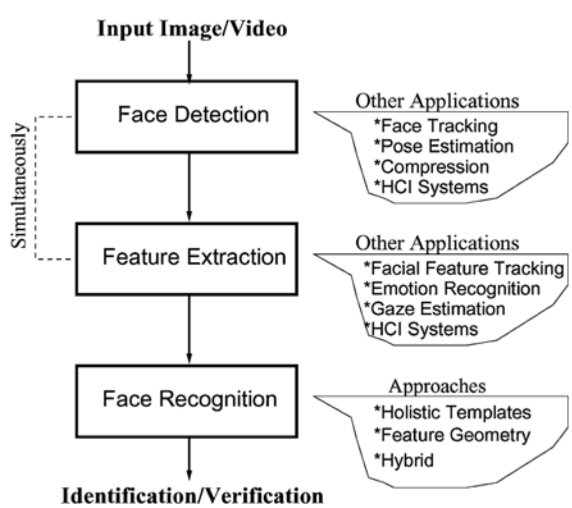

Figure 1. Schematic of a generic face recognition system

In the design of face recognition systems, at least three scenarios are to be kept in mind. These are:

- **Verification**: A recognition system determines if the person in a face image and matches a claimed identity.
- **Identification**: A recognition system determines the identity of a person in a face image.

- **Watch list**: A recognition system first determines if the person in a face image is on a watch list, and, if yes, then identifies the individual.

Figure 2 illustrates the above three tasks. The difficulty of the identification and watch list scenarios depends on the size of the database or watch list.

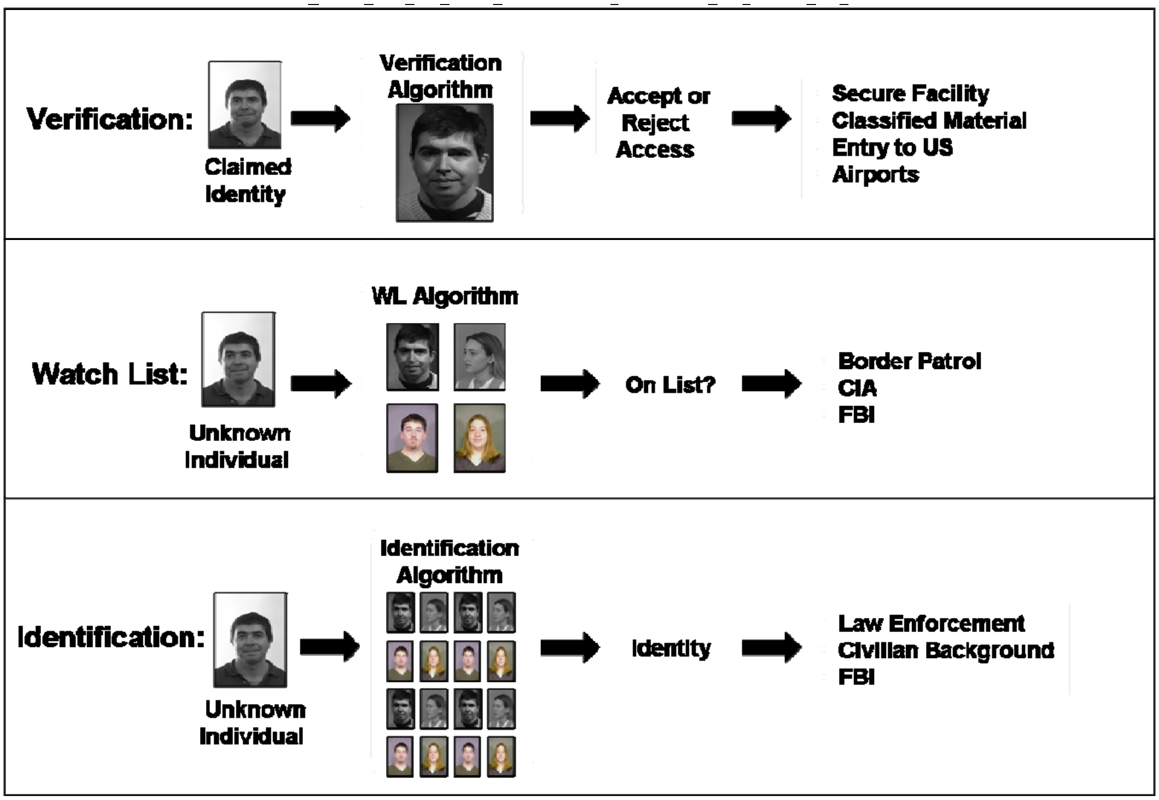

Figure 2. Illustration of three face recognition tasks: verification, identification, and watch list.

2. WHY FACE RECOGNITION IS HARD

Images of human faces undergo many changes due to acquisition conditions and natural aging. Acquisition conditions refer to the pose of the face with respect to the camera, illumination conditions, facial expressions and the number of pixels in the face region. Additional variations may be caused by disguises, occlusions (due to sun glasses, baseball hats, etc) and gain/loss of weight and facial hair. As part of aging one may undergo weight gain or loss, thus adding another dimension to the variations in human faces. Although the person is the same, the range of faces images can be very large. The range of faces of an individual is illustrated in Figure 3. The challenge of face recognition task is to be able to recognize a person in the presence of all these variations.

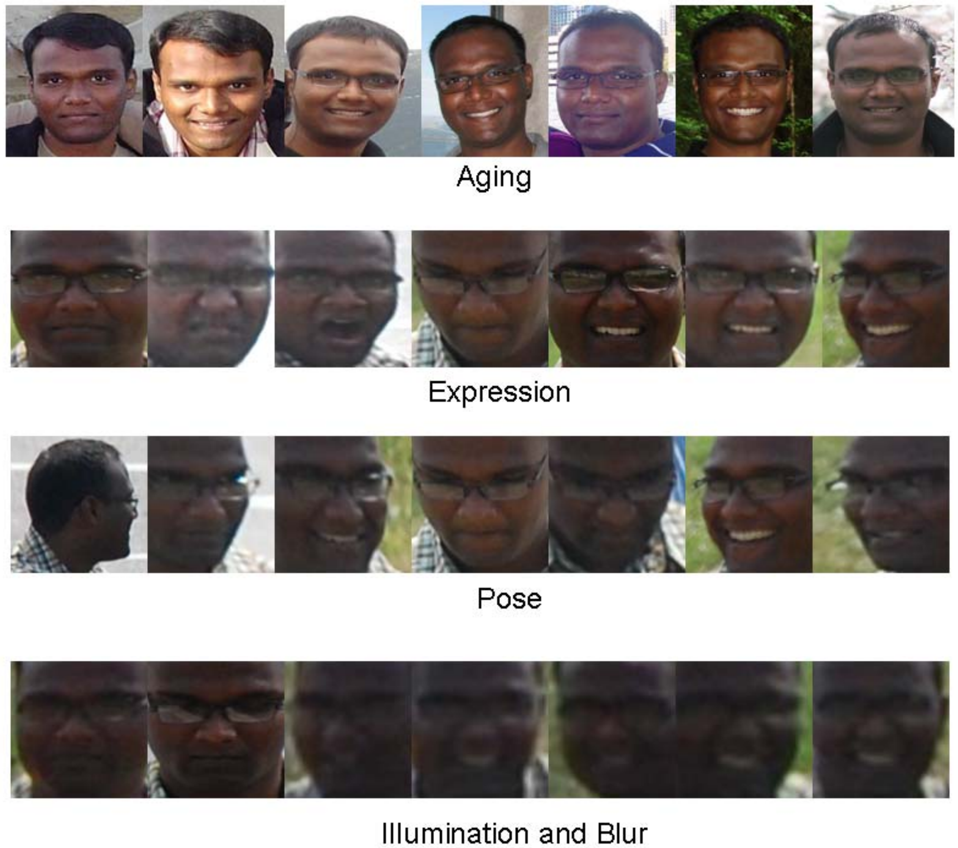

Figure 3. Range faces of the same individual due to variations in aging (first row—from youngest to oldest), expression (second row), pose (third row), and illumination and blur (bottom row).

3. HOW HUMANS PERCEIVE FACES

Since humans possess impressive skills for recognizing faces, it is worthwhile for designers of face recognition system to be cognizant of the factors that affect human perception of faces. This area has been widely studied for at least three decades. The readers are referred to a review paper that recently appeared [1]. From Sinha et al [1], we summarize key findings related to human face recognition organized into five categories.

Recognition as a function of available spatial resolution

- Humans can recognize familiar faces in very low-resolution images.
- The ability to tolerate degradations increases with familiarity.
- High-frequency information by itself is insufficient for good face recognition performance.

The nature of processing: Piecemeal versus holistic

- Facial features are processed holistically.
- Of the different facial features, eyebrows are among the most important for recognition.
- The important configural relationships appear to be independent across the width and height dimensions.

The nature of cues used: Pigmentation, shape and motion

- Face-shape appears to be encoded in a slightly caricatured manner.
- Prolonged face viewing can lead to high level after effects, which suggest prototype-based encoding.
- Pigmentation cues are at least as important as shape cues.
- Color cues play a significant role, especially when shape cues are degraded.
- Contrast polarity inversion dramatically impairs recognition performance, possibly due to compromised ability to use pigmentation cues.
- Illumination changes influence generalization.
- View-generalization appears to be mediated by temporal association.
- Motion of faces appears to facilitate subsequent recognition.

Developmental progression

- The visual system starts with a rudimentary preference for face-like patterns.
- The visual system progresses from a piecemeal to a holistic strategy over the first several years of life.

Neural underpinnings

- The human visual system appears to devote specialized neural resources for face perception.
- Latency of responses to faces in infero temporal (IT) cortex is about 120 ms, suggesting a largely feed forward computation.
- Facial identity and expression might be processed by separate systems.

4. REVIEW of STATE of The ART

4.1 Face Detection

The first step in any automatic face recognition systems is the detection of faces in images. After a face has been detected, the task of feature extraction is to obtain features that are fed into a face classification system. Depending on the type of classification system, features can be local features such as lines or fiducial points, or facial features such as eyes, nose, and mouth. Face detection may also employ features, in which case features are extracted simultaneously with face detection. An excellent survey of face detection algorithms developed before 2000 is given in [5].

One of the most popular and robust face detection algorithms is the one designed by Viola and Jones. Viola and Jones [6] introduced a machine learning approach for object detection by learning a strong classifier through a weighted combination of several weak learners. For a two class problem with labeled training examples, a learning algorithm based on adaboost selects a small number of critical visual features that provide the best classification accuracy. To detect faces in still images, a set of Haar wavelet-like features was input to weak learners to capture variations in the appearance patterns of faces and non-faces. An attentional cascade of adaboost stages was then designed such that the potential non-face regions were rejected early in the process before focusing more on face-specific regions in an image. The weak learners obtained from each stage of the cascade were then combined to produce a final strong classifier. An example of a typical face and feature detection algorithm is given in Figure 4.

Figure 4. An example of face detection and facial feature extraction [Taken from Moon et al, IEEE Transactions on Image Processing, Vol., pp., Nov. 2002]

4.2 Still-face recognition

The early years Face recognition research was energized in the late eighties and early nineties by the use of subspace methods such as the principal component analysis (PCA), Linear Discriminant Analysis (LDA) and a structural approach called elastic graph matching (EGM) matching. Since then, numerous researchers have extended these three types of algorithms have appeared in the literature. In the FERET evaluation of face recognition algorithms conducted in late 1996 and early 1997, it was found [7] that algorithms derived from a probabilistic subspace analysis, LDA and EGM performed well. The same experiments showed that when faces separated by up to 18 months were given to the recognition algorithms, the performance was poor. Starting in 2000, scientists and engineers from the National Institute of Standards and Technology (NIST) conducted the series of Face Recognition Vendor Tests (FRVT), See Table 1 for brief descriptions of these tests.

Table 1 How well do face recognition algorithms work?
A brief history of Face Recognition Vendor Tests.

Test	Description	Conclusions
FRVT 2000	FRVT 2000 was a technology evaluation that used the Sep96 evaluation protocol, but was significantly more demanding than the Sep96 FERET evaluation. Participation in FRVT 2000 was restricted to COTS systems A greater variety of imagery was used in FRVT 2000. FRVT 2000 reported results in eight general categories: compression, distance,	(a). Probe images compressed using JPEG up to 40:1 did not reduce recognition rates.(b) The evaluation results showed that pose does not significantly affect performance up to ±25◦, but that performance is significantly affected when the pose angle reaches ±40◦. (c) The indoor change of lighting did not significantly affect performance, but moving from indoor to outdoor lighting significantly affected performance (d) On the same images

	expression, media, illumination, pose, resolution, and temporal. There was no common gallery across all eight categories. .	in the FERET evaluation, performance improvements were on face images of a person were taken images at least a year apart. (e) Finally, the evaluations also showed that future areas of interest continue to be pose and illumination variations, and when faces images of a person are taken at least a year apart. These conclusions were taken from http://www.frvt.org. .
FRVT 2002	The primary objective was to provide performance measures for assessing the ability of automatic face recognition systems to meet real-world requirements. Ten participants were evaluated. Real-world performance statistics for verification and identification on a very large data set were computed.	(a) Indoor performance has improved since FRVT 2000. (b) Performance decreases approximately linearly with elapsed time. (c) Better systems are not sensitive to indoor lighting. (d) 3D morphable models improve performance. (e) Males are easier to recognize than females. (f) Older subjects are easier to recognize than younger subjects. (g) Finally, outdoor recognition performance needs improvement. These conclusions were taken from http://www.frvt.org.
FRVT 2006	The primary objective was to evaluate 3D and still image-based face recognition algorithms. The evaluations were organized along three experiments. The first experiment compared two still images taken with studio lighting. The second matched 3D face data using shape and texture information. The third compared a still image face image taken under studio lighting to still face images taken in hallways and atriums.	(a) It was found that two orders of magnitude improvement in recognition performance was obtained since 1993. (b) The recognition improvement was one order since 2002. (c) On comparably controlled acquisition conditions, the performance of iris and face-based recognition performance was comparable on the FRVT 2006 data. (d) The performance of still image and 3D face-based methods was comparable. (e) Under some conditions, computers can recognize faces better than humans. (f) Finally, illumination and resolution do matter in achieving high recognition rates. These conclusions were abstracted from Phillips et al, "FRVT 2006 and ICE 2006 Large Scale Results," *IEEE Trans. Patt. Anal. and Mach. Intel.*, (In press).

The Pose, illumination and Expression (PIE) problem Several researchers have addressed pose variation in face recognition. Some of the earlier attempts include extending the eigenface approach by building separate eigenspaces that capture information from different viewing directions, compensating for pose variation by building a 3D model and generating 2D representations for multiple poses. To handle pose and illumination variations, a 3D morphable face model was proposed in [8], where the shape and texture of each face is represented as a linear combination of a set of 3D face exemplars and the parameters are estimated by fitting a morphable model to the input image. By far the most impressive face synthesis results were reported in [8] accompanied by very high recognition rates. Many extensions of the 3D morphable model approach have appeared in the literature with different degrees of successes. Most of the 3D morphable model approaches are computationally intensive and often need a small number of features to be manually selected.

Parallel to the development of pose normalization methods, approaches to illumination normalization have engaged the attention of computer vision researchers. Earlier attempts at reducing the effects of illumination include dropping the first few eigenvalues of the principal component expansion, using the gradient directions as features or building a subspace representation known as the illumination cone to capture the images that a convex Lambertian object can produce under all lighting conditions. Low-dimensional spherical harmonics representations were also found effective for face recognition under lighting variations. Extensions to 3D morphable model-based approach to derive lighting invariant representations have also been proposed. Other efforts include computing a self-quotient image by dividing a face image by a smoothed version of the image, leading to insensitivity to lighting variation, a generalized photometric stereo algorithm that allows for within-class shape variation. More recently, a non-stationary stochastic filtering algorithm for estimating illumination-insensitive albedo maps for face recognition has been developed. An example of estimating illumination-free albedo maps and 3D models from a single image are shown in Figure 5. The general consensus is that although these methods have produced much better results than the traditional subspace methods for recognizing faces with illumination variations, they have all been evaluated on controlled data sets as the Yale B data set or the PIE data set collected at the Carnegie Mellon University. Designing methods that are robust to illumination variations in uncontrolled situations is an open problem.

Facial expression analysis and recognition has been extensively studied in the context of human-computer interactions. As observed in Section 3, facial identity and expression might be processed by separate systems. Although many techniques for automatic recognition of expressions are available, they are effective for macro expressions such as happy, anger, surprise, fear, etc. Analysis and recognition of micro-expressions is an active research area and has penetrated the popular culture by way of the television show **Lie to Me!** For more detailed discussions on facial expressions, the reader is referred to [9].

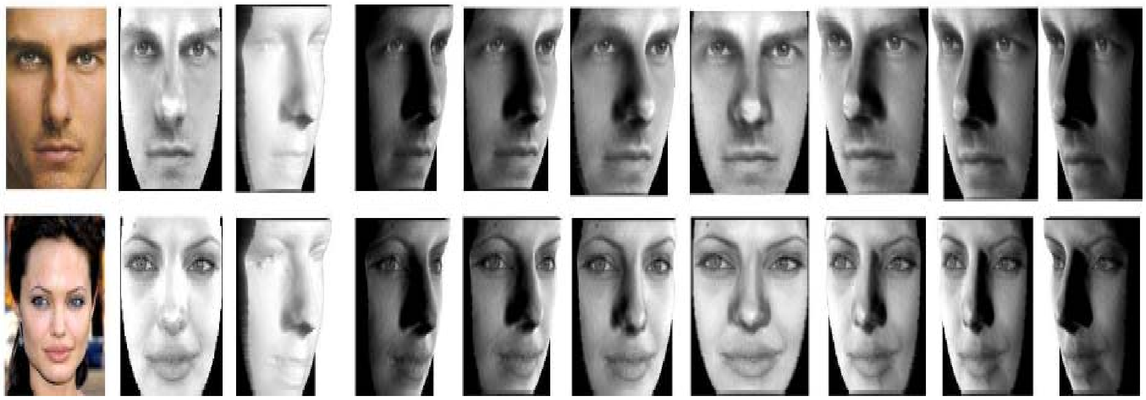

Figure 5. Examples of recovering illumination-free albedo maps and 3D models from an image downloaded form the internet. In each row, the left most image is the image downloaded from the internet. The next two images are the reconstructed 3D models in two poses. The next set of images are generated by synthesizing new images from the 3D models corresponding to different poses.

The reader is referred to a survey paper [3] for a more detailed account of face recognition research reported until end of 2002.

Table 2 It's all about eyebrows!

As discussed in [1], of the different facial features, eyebrows are among the most important for recognition. Sadr et al. have presented evidence suggesting that the eyebrows might not only be important features, but that they might well be among the most important, comparable to the eyes. How might one reasonably explain the perceptual significance of eyebrows in face recognition? There are several possibilities. First, eyebrows appear to be very important for conveying emotions and other nonverbal signals. Since the visual system may already be biased to attend to the eyebrows in order to detect and interpret such signals, it may be that this bias also extends to the task of facial identification. Second, for a number of reasons, eyebrows may serve as a very stable facial feature. Because they tend to be relatively high-contrast and large facial features, eyebrows can survive substantial image degradations. Also, since eyebrows sit atop a convexity (the brow ridge separating the forehead and orbit), as compared to some other parts of the face, they may be less susceptible to shadow and illumination changes.

J. Sadr, I. Jarudi and P. Sinha, P, "The Role of Eyebrows in Face Recognition," Perception, 32, 285-293.

4.3 Face recognition across aging

Face recognition across aging is most challenging in that it has to address all other variates as well. Pose, expression and illumination changes are bound to happen for two images of a person taken years apart. In addition to this, textural properties of the skin can be different as well (makeup, spectacles, weight loss/gain, hair loss, etc). According to [10], the facial changes that occur due to aging are influenced by numerous environmental factors like solar radiation, smoking, drug usage, stress level, etc. The different biological and environmental factors can either delay or expedite the process of aging. Aging results in changes in both the hard and soft facial tissue of an individual. Loss of tissue elasticity and facial volume and alteration in skin texture are some of the other changes with aging. But although the manner of aging is highly unpredictable, there is a sequence of changes that appears to adhere to a basic progressive pattern across time. Drifts in facial landmarks appear to reasonably characterize the shape variations associate with aging, especially in ages 2-18. For older subjects, variations in facial texture appear to dominate variations in shape. Contributions to face morphological studies have come from both psychophysics and computer vision researchers. Methods from psychophysics include deriving cardioidal strain transformations and their extensions, variations in shape and degree of skin wrinkling and exaggeration or a de-emphasis of facial creases. Computer vision researchers have proposed subspace-based, model-based and machine learning approaches for face recognition across aging.. See Table 3 for a brief summary of recent work in this area. Figure 6 presents examples of synthesizing aged faces in the age group 2-18 years.

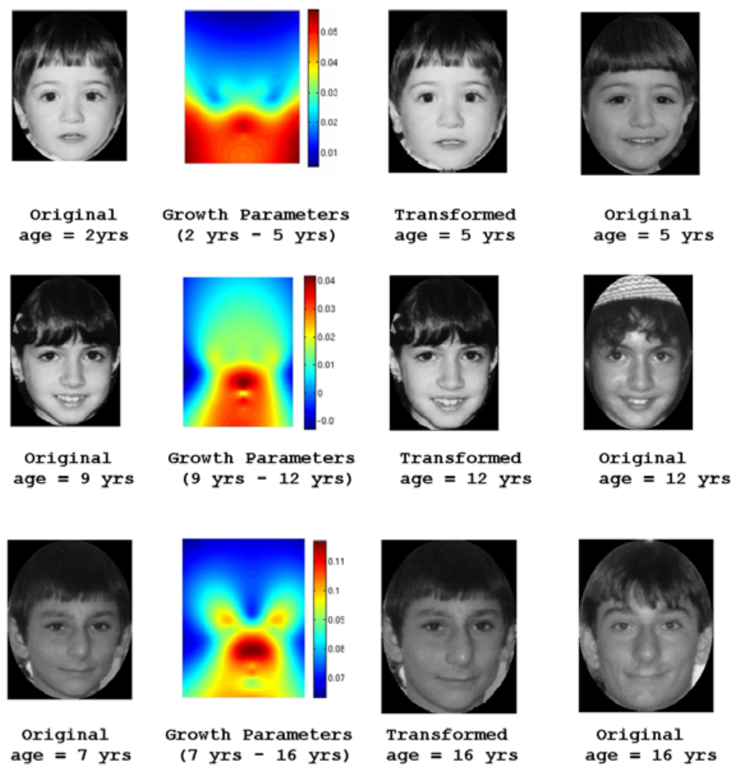

Figure 6. Some appearance prediction results derived using the craniofacial growth model discussed in [11]. The first column is original images of children. The second column is the estimation of growth changes. The third column is the algorithm estimated aging of a child, and the fourth column is an image of the child at the transformed aged.

Table 3 Examples of face recognition across aging (See [11] for information on cited references in this table)

Reference	Summary of contributions
Shaw et al	Sought to identify mathematical transformations that help characterize the facial growth event. They discovered two transformations that could be applied on the outer contour of faces in the 'profile view' namely, cardioidal strain, which stretches the face downward and outward and affine shear, which when applied in the right proportion, introduces a protrusion in the jaw and a backward slant in the forehead
Pittinger and Shaw	Revisited the above approach and investigated the relative importance of three force configurations namely, the shear forces, the strain forces and the radial forces in inducing facial growth.
Todd et al	Proposed the hydrostatic model, also called as the 'revised' cardioidal strain transformation model to characterize facial growth. Drawing analogies between human head growth and the modeling of a fluid-filled spherical object with pressure, they performed a hydrostatic analysis of the effects of gravity on a growing head.
Mark et al	Hypothesized that the perceptual information associated with any recognizable style of change were contained in the geometric invariants associated with the event. Three geometric invariants were identified in relevance to facial growth, namely, the angular coordinates of features being preserved; bilateral symmetry about the vertical axis being preserved and continuity of all contours and their directions of curvature being preserved.
Mark and Todd	Extended the 2D cardioidal strain transformation model into 3D and demonstrated its effectiveness in characterizing facial growth in 3D
Bruce et al.	Observed that a subject's sensitivity to cardioidal strain related changes in 3D faces were comparable, when viewed in pairs of face profiles or pairs of 3/4 faces or pairs of mixed profiles
O'Toole et al.	Applied a standard facial caricaturing algorithm on 3D faces. Noted that an exaggeration or a de-emphasis of facial creases, increased or decreased the perceived age of faces respectively.

4.4 Video-based face recognition

Video-based face recognition (VFR) is the technique of establishing the identity of one or multiple persons present in a video, based on their facial characteristics. Given the input face video, a typical VFR approach combines the temporal characteristics of facial motion with appearance changes for recognition. This often involves temporal characterization of faces for recognition, building 3D model or a superresolution image of the face, or simply learning the appearance variations from the multiple video frames. The ability to generalize across pose, illumination, expression, etc. depends on the choice of combination. Video-based face recognition is particularly useful in surveillance scenarios in which it may not be possible to capture a single good frame as required by most still image based methods.

A typical VFR system operates by acquiring video feeds from one or multiple cameras, tracking and segmenting faces from the input feed(s), extracting representations to characterize the identity of the face(s) in the video, and then comparing them with the enrolled representations of subjects in the database. This constitutes the test phase of the system. During the enrollment (or training) phase, a similar sequence of steps is followed using one or multiple video feeds per identity and the corresponding composite representations are stored in the database. VFR approaches differ in the representation that is used to characterize the moving faces. An ideal VFR system performs these operations automatically without any human intervention.

Effective utilization/fusion of the information (both spatial and temporal) present in a video to achieve better generalization (for each subject) and discriminability (across different subjects) for improved identification is one of the biggest challenges faced by a
VFR system. The fusion schemes can range from simple selection of good frames (which are then used for recognition in a still-image based recognition framework) to estimation of the full 3D structure of a face which can then be used to generalize across illumination, pose, etc. The choice may depend primarily on the operational requirements of the system. For example, in a surveillance setting, the resolution of the faces may be too small for reliable shape estimation. The choice also limits the recognition capability of the system. A simple good frame selection scheme will not have the capability to generalize appearance across pose variations and thus requires the test video to have some pose overlap with the gallery videos. Effective modeling of subject-specific facial characteristics from video data can only be achieved if the changes in facial appearance during the course of the video are appropriately attributed to different factors like pose changes, lighting, expression variations, etc. Unlike still image based scenarios, these variations are inherent in a VFR setting and must be accounted for to reap the benefits of extra information provided by the video data. In addition, due to the nature of the input data, VFR is often addressed in conjunction with tracking problem which is a challenging problem by itself. In fact, more often than not, tracking accuracy depends on the knowledge of reliable appearance model (depends on the identity provided by the recognition module) while recognition result is dependent on the localization accuracy of the face region in input video.

Existing VFR systems have been designed using a simultaneous tracking and recognition approach, a 2D feature graph matching across the temporal axis, a 3D model-based approach, hidden Markov models and probabilistic appearance manifolds. Table 4 gives a brief summary of existing methods.

Table 4 Examples of video-based face recognition systems

Algorithm	**Short description**	**Experimental description**
Probabilistic recognition of human faces from video Zhou, et al, CVIU, 2003.	Simultaneous tracking-and recognition using a dynamic state space model and sequential importance sampling	Private: 12 subjects, NIST: 30 subjects, MoBo : 25 subjects

VFR using probabilistic appearance manifolds. Lee et al, CVIU 2005. Turaga, et al, CVPR 2008	Face modeled using a low-dimensional appearance manifold, approximated by piecewise linear subspaces, and special manifolds.	Honda-UCSD dataset: 20 subjects (52 videos)
VFR through tracking facial features, Li, et al, JOSA 2001.	Tracks facial features defined on a grid with Gabor attributes using SIS algorithm	Li dataset: 19 subjects (2 sequences each
VFR using adaptive hidden Markov models Liu and Chen, CVPR 2003.	Statistics of training videos, and their temporal dynamics learnt by an HMM.	Private: 12 subjects, Mobo [8]: 25 subjects

5. FUTURE RESEARCH DIRECTIONS

5.1 Human perception of faces

Although we now have an impressive body of experimental literature on human perception of faces, several fundamental issues are still unresolved. We list a few of the key ones here:

1. Precisely what configural information is important for recognition? Most studies have tended to gloss over this question by focusing on the distinction between *configural* and feature-based approaches to face recognition. It is clear that something about the overall gestalt of the face is important. What is not known is exactly how to operationalize this general notion of face gestalt. Which facial measurements contribute to this encoding?
2. How does familiarity change facial representations? Human face recognition processes can tolerate greater degradations in the images of people we are familiar with relative to those that we have only a casual acquaintance with. This suggests that the internal facial representation undergoes important changes with increasing familiarity. What is the nature of these changes? Does encoding progress from being more piecemeal to more holistic with greater experience? How do these changes confer greater robustness to transformations?
3. What is the role of top-down expectations on recognition? As mentioned in section 3, the latency of response of face selective neurons in the primate infero-temporal cortex is just a little over 100 ms. Given conventional ideas of rate-coding, this low latency suggests that face processing might be largely feed-forward in nature. If so, how can prior expectations influence identity computation? Furthermore, under what conditions can top-down influences usefully contribute to face recognition?

Answering these questions promises not only to shed light on brain mechanisms of face recognition, it will also provide clues for the development of more effective strategies and representations suitable for deployment in computer-vision based systems.

Are Computers Better than Humans?

Phillips et al showed that computers can out perform humans on frontal still face images across changes in illumination. How general is this result? Humans are very good at recognizing familiar faces and poor at recognizing unfamiliar faces. Because of our ability to recognize familiar faces, we over estimate our skill at recognizing unfamiliar faces. Even for recognizing unfamiliar faces, humans are the most robust face recognition systems available. Humans can adjust for combinations of changes in pose, illumination, blur, and resolution significantly better than computers. In low-resolution video, humans intrinsically integrate temporal and body features that leading-edge research is only now starting to address. Recent work has shown that fusing computers and humans can lead to near perfect recognition (O'Toole et al, IEEE:TSMCA 2007).

P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, M. Sharpe, "FRVT 2006 and ICE 2006 Large Scale Results," *IEEE Trans. Pattern Analysis and Machine Intelligence*, (In press)

5.2 Remote face recognition

Most of the existing face recognition algorithms and systems are effective when the face images are within few tens of meters form the camera. Extending the distance at which face recognition systems can be effective is a new thrust with applications in surveillance. In the remote acquisition scenario, the face images are often blurred, may not always have sufficient number of pixels on faces and may have significant pose and illumination variations as well as occlusion. Figure 7 shows images of several subjects acquired at various distances from the camera. In the remote scenario, acquiring face signatures that are of sufficient quality to be fed into recognition engines is itself a challenge. This is especially true when the sensor and the subjects are moving. In this case, one needs to stabilize the videos and robustly track the moving face before it can be recognized.

.

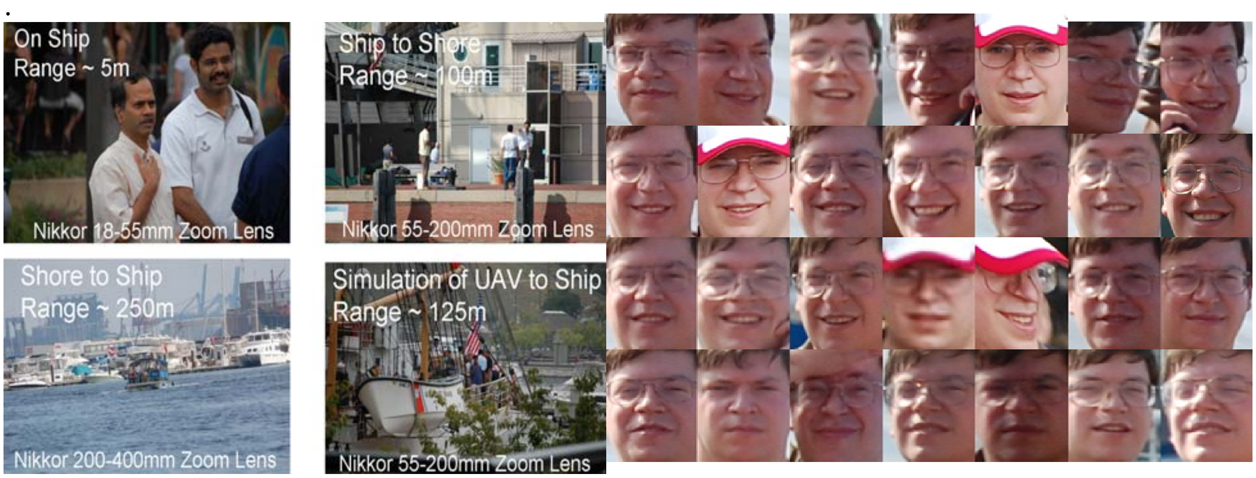

Figure 7: Examples of face images acquired at different distances from the camera.

5.3 Video-based face recognition

Robust approaches that exploit video sequences are needed for many applications such as maritime security and other such security needs. Video-based face recognition has received attention over the past nine years. In the early stages of development, VFR research had to cope up with lack of video data for evaluation. Under the NIST Multi-biometrics Grand Challenge (MBGC) program, several hundred of video sequences have been made available for evaluating single-gallery to video matching problem and video-to-video matching problem. This is a positive development for this area. To be effective, the following problems have to be addressed. Real-time tracking and pose normalization of moving faces, illumination normalization, compensation for low-resolution face images via superresolution techniques and simultaneous tracking and recognition. Algorithms that can accommodate multiple gallery images or gallery and probe video sequences have to be developed.

5.4 Face recognition in a camera network

Multi-camera networks are becoming increasingly common for wide-area surveillance problems. Having multiple face images acquired by a camera network help build more robust descriptions of faces as increases the chance of the person being in a favorable pose (frontal or near frontal). However, to use the multi-view information, we need to estimate the pose of the subject's head. This could be done explicitly by computing the actual pose of the subject to a reasonable approximation, or implicitly, by using a view selection algorithm. But solving for the pose of the subject's head is difficult, especially when the resolution of the images is low and the calibration of cameras (both external and internal) is not sufficiently precise to allow robust multi-view fusion. This is true when the subjects are usually in the far field of cameras. In addition, many other problems such as multi-view tracking, appropriate representations for multi-view face images and multi-view recognition of faces need additional investigation. Another major issue to be resolved is whether the algorithms have to be centralized or distributed. Some examples of faces acquired using a multi-view camera network are shown in Figure 8.

Registration, registration and registration!

At a meeting held in winter of 2002 to discuss the future challenges in face recognition, Prof. Takeo Kanade claimed face alignment is a critical issue that should not be ignored. He based his assertion on the experiments done using the CMU PIE data set. Prof. Kanade's claim has been more than validated. This has also been noted by other researchers. In the mid nineties, Prof. Tomaso Poggio pointed out that if faces are not perfectly aligned, even the simplest of face recognition techniques, such as the principal component analysis is not valid! If faces are not registered well, illumination normalization methods based on pixel descriptions such as generalized photometric stereo, self-quotient image and shape from shading suffer. Registration of face alignment is critical for mitigating the effects of pose variations. Over the years, computer vision researchers have struggled with this problem and have suggested many techniques for face alignment using feature graphs, 3D morphable models, etc. Automatic registration of faces is still an open problem.

Figure 8. Faces of different subjects acquired in a camera network

5.5 Face Recognition in Web 2.0

Within the last year, face recognition modules have been added to Google's[1] Picasa, Facebook, and iPhoto. These modules are designed to recognize faces in a person's photo library or Facebook network. These modules allow for users to correct mislabeled faces. The feedback from users will lead to rapid identification of areas were automatic face recognition fails and hence research is needed. Many of the same issues in multi-view camera networks apply to face recognition in Web 2.0. However, there are unique aspects to this application; e.g., deriving algorithms that correctly label faces in overlapping networks of social contacts.

5.6 Face recognition across aging

Existing age estimation algorithms are effective for determining ages only within a few years. The synthesis of aged faces for subjects in the group 2-18 is largely determined by shape variations, while for adults, shape and texture variations come into play, with texture variations dominating the shape variations. In a recent work [11], brief discussions on many models for aging adults were presented. One of the modes comprises of a shape variation model and a texture variation model. Attributing facial shape variations during adulthood to the changing elastic properties of the underlying facial muscles, the shape variation model was formulated using physical models that characterize the functionalities of different facial muscles. Identifying facial muscles into one of three types namely, (i) linear muscles (ii) sheet muscles (iii) sphincter muscles, the authors of [11] proposed transformation models for each and modeled facial feature drifts as linear combinations of the drifts observed on the individual facial muscles. The texture variation model was designed specifically to characterize facial wrinkles in pre-designated facial regions such as the forehead, nasolabial region etc. Figure 9 illustrates the facial muscle configurations and further illustrates the pressure models that were adopted for each type of facial muscle. The synthesis approach needs to be validated for its effectiveness for face recognition across aging.

[1] The identification of any commercial product or trade name does not imply endorsement or recommendation by the authors or their institutions.

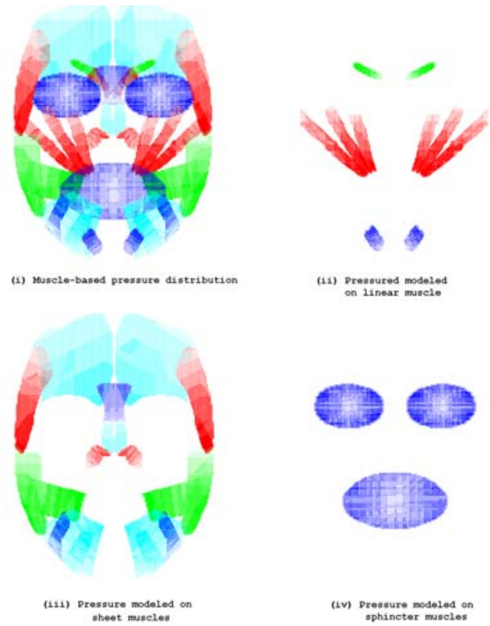

Figure 9. An illustration on the configuration of facial muscles and the proposed pressure models.

An alternative approach that does not rely on aging models is to simply ask the question if the two age-separated faces belong to the same individual. One needs to extract age-invariant features for this approach to be useful. Recent efforts that have taken the non-generative approach aim to derive metrics for measuring the cohesiveness of feature drifts of age-separated face images of the same subject and compare it with age-separated faces of different subjects. The problem of designing appropriate representations and decision rules for face recognition across aging is an open problem.

Face recognition: Does skin matter?

Skin color and surface roughness may enable rapid classification of a person into an ethnic and age group by exploiting skin pigmentation, translucency characteristics and patterns of surface bumps. These attributes help prune the set of candidate gallery matches to a probe.

Moles, freckles and scars are local skin attributes that are, if present and durable, extremely powerful recognition cues since there is very low likelihood that different people will have the exact same local attributes. Local skin irregularities have different levels of permanence. While markings such as skin moles and freckles are permanent, most scratches and red spots are transient. The challenge, however, is to cope with variations in skin appearance that stem from natural and imaging causes (e.g., rate of blood flow to the skin and illumination variations, respectively).

Skin roughness can be divided into local and global components. Forehead and crows' feet wrinkles are examples of local surface roughness. Global roughness may emerge due to age or health causes and typically covers large areas of the face (e.g., pimples). Skin surface roughness is typically visible only in medium and large resolution images. It has the advantage of appearing relatively robust to adverse factors such as facial expression, illumination and head pose.

We thank Dr. Yaser Yacoob for providing this column.

5.7 Face and other biometrics

In order to ensure robustness and face recognition algorithms have often worked in conjunction with fingerprint, iris, gait and voice recognition systems. This has led to the creation of a research area that is called as multi-modal or multi-biometrics systems. One of the main challenges in fusing biometrics algorithms and/or systems is to come up efficient and robust fusion results. This area has benefitted largely from the theory and design of multiple classifier systems. Although there are several examples of face/fingerprint, face/gait, face/voice and face/iris fusion, the area of multi-modal biometrics, where face is one of the biometrics signatures is in its infancy. More discussions on the design of multi-biometric systems may be found in [12].

6. SUMMARY and CONCLUSIONS

In this paper, we have presented a brief summary of the state of the art in face recognition research both from human and machine perception points of view. We have also presented a brief account of future challenges as we see. It is our opinion that research in face recognition is an exciting area for many years to come and will keep many scientists and engineers busy. In additions robust face recognition systems have many applications in homeland security, human computer interaction and many consumer applications.

7. REFERENCES

1. P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, " Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About", Proceedings of THE IEEE, Vol. 94, pp. 1948-1962, Nov. 2006.
2. W. Zhao, R. Chellappa, J. Phillips and A. Rosenfeld, “Face Recognition in Still and Video Images: A Literature Surrey”, ACM Computing Surveys, Vol. 35, pp. 399-458, Dec. 2003.
3. S. Li and A.K. Jain (eds.), *Handbook on Face Recognition*, Springer, 2005.
4. W. Zhao and R. Chellappa, (eds.), *Face Processing: Advanced Modeling and Methods,* Academic Press, 2006.
5. Ming-Hsuan Yang; Kriegman, D.J.; Ahuja, N. " Detecting faces in images: a survey", IEEE Trans. on Patt. Anal. And Mach. Intell., Vol. 24, pp. 34-58, Jan. 2002.
6. P. Viola and M.J. Jones, “Robust Real-Time Face Detection", International Journal of Computer Vision, Vol. 57, pp. 137–154, 2004.
7. Phillips, P.J.; Hyeonjoon Moon; Rizvi, S.A.; Rauss, P.J, "The FERET evaluation methodology for face-recognition algorithms", IEEE Trans. on Patt. Anal. And Mach. Intell., Volume 22, pp. 1090-1104, Oct. 2000.

8. Blanz, V.; Vetter, T., "Face recognition based on fitting a 3D morphable model", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol**.** 25, pp. 1063-1074, Sept. 2003.
9. P. Ekman and E.L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (Series in Affective Science),* Kindle Book, 2005.
10. M. Albert, K. Ricanek and E. Patterson, A review of the literature on the aging adult skull and face: Implications for forensic science research and applications, Journal of Forensic Science International, April 2007.
11. N. Ramanathan and R. Chellappa, "Computational Methods for Modeling Facial Aging: A Survey", Journal of Visual Languages and Computing, Vol. 20, pp. 131-144, May 2009. (Invited Paper).
12. A. Ross, K. Nandakumar and A.K. Jain, *Handbook of Multibiometrics,* Springer, 2006