FACE RECOGNITION BY COMPUTERS AND HUMANS

Rama Chellappa, *University of Maryland, College Park*
Pawan Sinha, *Massachusetts Institute of Technology*
P. Jonathon Phillips, *National Institute of Standards and Technology*

Research into face recognition offers research opportunities that will challenge scientists and engineers for years to come. Robust face-recognition systems, for example, can provide applications in homeland security, human-computer interaction, and many consumer applications.

In most situations, face recognition is an effortless task for humans. But is this true for computers? That 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 already been deployed at ports of entry at international airports in Australia and Portugal.

Studies of how humans perceive faces have generated many interesting findings that can be used to help design practical systems.[1] Besides applications related to identification and verification—such as access control, law enforcement, ID and licensing, and surveillance—face recognition has also proven useful in applications such as human-computer interaction, virtual reality, database retrieval, multimedia, and computer entertainment.[2-4]

Figure 1 shows a schematic of a general face-recognition system, which consists of three major modules: face detection, feature extraction, and face recognition. As in any pattern recognition problem, variations in patterns from illumination, pose, expressions, and so on are handled either by making the features invariant or robust to these transformations in the feature extraction stage, or by designating rules that account for these transformations in the recognition stage.

In the design of face-recognition systems, at least three tasks must be kept in mind:

- *Verification*. A recognition system determines if the person pictured in a face image matches a claimed identity.
- *Identification*. A recognition system determines a person's identity in a face image.
- *Watch list*. A recognition system determines if the person in a face image appears on a watch list and, if so, identifies that individual.

 Published by the IEEE Computer Society

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

WHY FACE RECOGNITION IS HARD

Acquisition conditions—the face's pose with respect to the camera, illumination, facial expressions, and number of pixels in the face region—and natural aging cause human face images to undergo many changes. Additional variations can be caused by disguises, occlusions from items such as sunglasses or a hat, and facial hair. Aging also causes some people to experience weight gain or loss, thus adding another dimension to human face variations. Even when focusing on a single subject, the range of face images can be extensive, as Figure 3 shows. The challenge of face recognition is to identify a person in the presence of all these variations.

PERCEIVING FACES

Since humans possess impressive skills for recognizing faces, face-recognition system designers should be aware of the factors affecting human-face perception. This area, widely studied for at least three decades,[1] has provided key findings and can be organized into five categories.

Recognition

Humans can recognize familiar faces as a function of available spatial resolution, even in very-low-resolution images. The ability to tolerate degradations increases with familiarity, but high-frequency information by itself is insufficient for good face-recognition performance.

Piecemeal versus holistic

While the nature of face-recognition processing is piecemeal rather than holistic, facial features are processed holistically. Of the different facial features, eyebrows are among the most important for recognition. The configural relationships appear to be independent across the width and height dimensions.

The nature of cues

Face shape appears to be encoded in a caricatured way that focuses on pigmentation, shape, and motion. Prolonged face viewing can lead to high-level af-

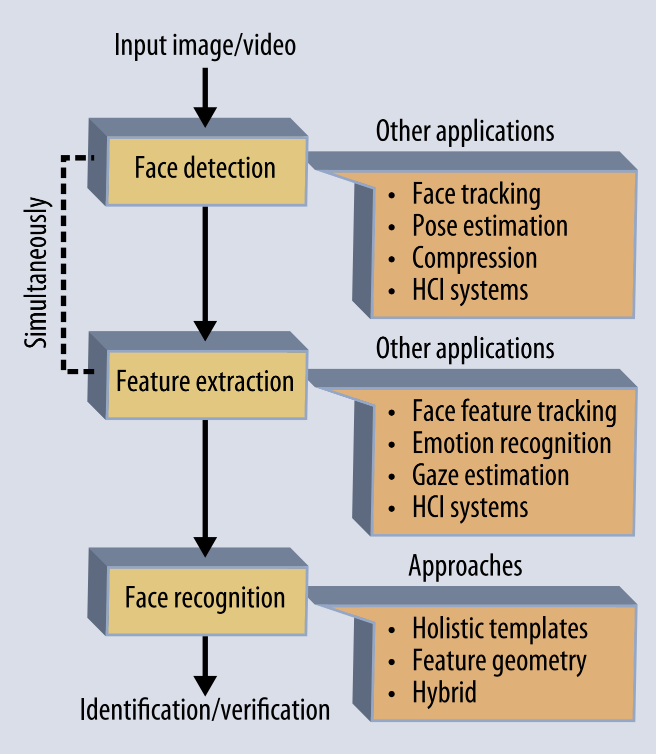

Figure 1. Schematic of a generic face-recognition system, which consists of three major modules: face detection, feature extraction, and face recognition.

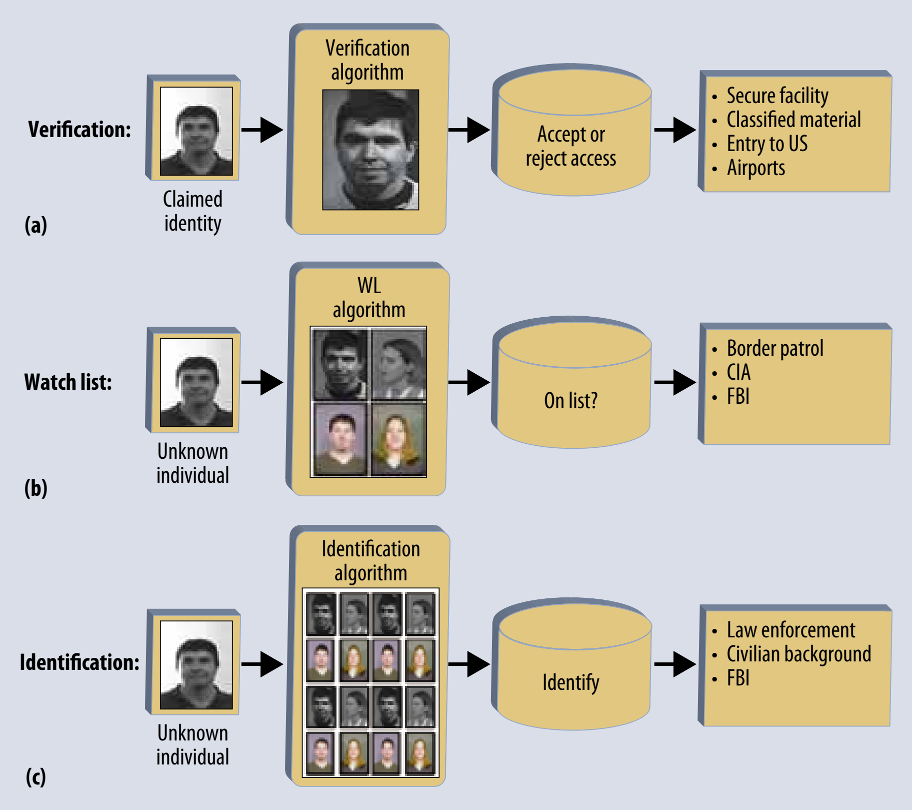

Figure 2. This diagram shows the structure of three face-recognition tasks. These tasks are verification, watch list, and identification.

ARE COMPUTERS BETTER THAN HUMANS?

Recent research shows that computers can outperform humans on frontal still-face images across changes in illumination.[1] How general is this result? Humans excel at recognizing familiar faces, but we overestimate our skill at recognizing unfamiliar ones. Yet even when recognizing unfamiliar faces, humans have the most robust face-recognition systems available. They can adjust for combinations of changes in pose, illumination, blur, and resolution significantly better than computers can.

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 (A.J. O'Toole et al., "Fusing Face-Verification Algorithms and Humans," *IEEE Trans. Systems, Man, and Cybernetics*, 2007, pp. 1149-1155).

Reference

1. P.J. Phillips et al., "FRVT 2006 and ICE 2006 Large Scale Results," *IEEE Trans. Pattern Analysis and Machine Intelligence*, forthcoming; DOI 10.1109/TPAMI.2009.59.

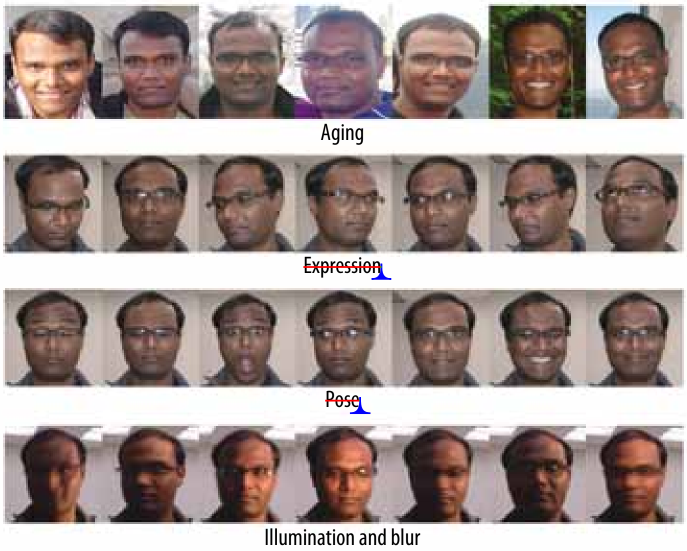

Figure 3. Range of faces for the same individual due to variations in aging: first row, from youngest to oldest; second row, expression; third row, pose; bottom row, illumination and blur.

tereffects that can be as straightforward as a face distorted in the opposite manner as the adapting face, or as complex as an antiface with a specific identity and no discernible distortions, suggesting the possibility of prototype-based encoding. Pigmentation cues are at least as important as shape cues.

Color cues play a significant role as well—especially when shape cues degrade—while contrast polarity inversion dramatically impairs recognition performance, possibly due to a compromised ability to use pigmentation cues. Illumination changes influence generalization, which appears to be mediated by temporal association. Face motion appears to facilitate subsequent recognition.

Developmental progression

The visual system starts with a rudimentary preference for face-like patterns. The system then 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 to face perception. The latency of responses to faces in the inferotemporal cortex is about 120 ms, suggesting a largely feed-forward computation. Facial identity and expression might be processed by separate systems.

STATE-OF-THE-ART REVIEW

The first step in any automatic face-recognition system is *face detection* in images.

Face detection

Once a face has been detected, feature extraction obtains information that can be fed into a face-classification system. Depending on a classification system's type, features can be local, such as lines or fiducial points, or facial landmarks, such as eyes, nose, and mouth.[5]

Paul Viola and Michael Jones[6] designed one of the most popular and robust face-detection algorithms. They 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.

Figure 4 shows an example of a typical face- and feature-detection algorithm.

The early years

In the late 1980s and early 1990s the use of still-face-recognition subspace methods such as *principal component analysis* (PCA), *linear discriminant analysis* (LDA), and a structural approach called *elastic graph matching* (EGM) energized face-recognition research. Since then, many researchers have extended these three algorithm types.

In the Face Recognition Technology (FERET) evaluation of face-recognition algorithms conducted in late 1996 and early 1997,[7] the best performers were algorithms derived from a probabilistic subspace analysis, LDA, and EGM approaches. The most difficult FERET experiment was recognizing faces of a person from images taken at least 18 months apart. Table 1 provides brief descriptions of the series of *face-recognition vendor tests* (FRVTs) conducted by the National Institute of Standards and Technology (NIST) starting in 2000.

Pose, illumination, and expression

Researchers have addressed face recognition across changes in pose, illumination, and expression (PIE). Earlier attempts included 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, developers proposed a 3D morphable face model[8] in which a linear combination of a set of 3D face exemplars provides the parameters for the shape and texture, and the parameters are estimated by fitting a morphable model to the input image.

Figure 4. Example of face detection and facial feature extraction. (Image courtesy of Hankyu Moon and colleagues, *IEEE Trans. Image Processing*, Nov. 2002, pp. 1209-1227.)

Table 1. Summary of FRVT evaluations.

Test	Description	Conclusions
FRVT 2000	The Face Recognition Vendor Test 2000 technology evaluation used the Sep96 FERET evaluation protocol, but was significantly more demanding than the Sep96 FERET evaluation. Participation in FRVT 2000 was restricted to COTS systems. A variety of imagery was used in FRVT 2000, which reported results in eight general categories: compression, distance, expression, media, illumination, pose, resolution, and temporal. There were no common findings across all eight categories.	(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 FERET experiment where face images of a person were taken at least 18 months apart, performance improved substantially. (e) Recognition across changes in pose and illumination continue to be future areas of interest; along with face images of a person taken at least a year apart. These conclusions were taken from 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 were reported on a very large dataset of 121,589 face images of 37,437 people.	(a) Indoor performance had 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) Outdoor recognition performance needs improvement. These conclusions were taken from 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 face image taken under studio lighting to still-face images taken in hallways and atriums.	(a) A two-orders-of-magnitude improvement in recognition performance has been obtained since 1993. (b) The recognition improvement was one order since 2002. (c) On comparably controlled acquisition conditions, the performance of single iris- and face-based recognition performance was comparable with the FRVT 2006 data. (d) The performance of still-image-based and 3D face-based methods was comparable. (e) Under some conditions, computers can recognize faces better than humans can. (f) 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," to be published in *IEEE Trans. Pattern Analysis and Machine Intelligence*; DOI 10.1109/TPAMI.2009.59.

REGISTRATION, REGISTRATION, REGISTRATION

At a meeting held in 2002 to discuss the future challenges in face recognition, Takeo Kanade ~~asserted~~ that face alignment poses a critical issue that should not be ignored. He based this ~~assumption~~ on experiments done using the CMU PIE dataset (http://facealignment.ius.cs.cmu.edu/alignment/webdemo.html).

Kanade's claim has been validated and also has been noted by other researchers. If faces are not registered well, illumination-normalization methods based on pixel descriptions such as generalized photometric stereo, self-quotient imaging, shape-from-shading, and 3D morphable models 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 suggested many techniques for face alignment using feature graphs, 3D morphable models, and related techniques. Yet automatic registration of faces remains an open problem.

Consistently, 3D morphable model-based approaches yielded high recognition rates when recognizing nonfrontal faces. Many extensions of these approaches have met with different degrees of success. Most of the model-based approaches are computationally intensive and often require manually selecting a small number of features.

Parallel to the development of 3D morphable model-based methods, approaches to illumination normalization have engaged the attention of computer vision researchers. Earlier attempts at reducing illumination effects included 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 of a convex Lambertian object.

Low-dimensional spherical harmonics representations were also found effective for face recognition under lighting variations. Extensions to a 3D morphable model-based approach to derive lighting-invariant representations have also been proposed. Other efforts included computing a self-quotient image by dividing a face image with 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, researchers have developed a nonstationary stochastic filtering algorithm for estimating illumination-insensitive face-recognition albedo maps. Figure 5 shows examples of estimating these maps and 3D models from a single image. The general consensus is that while 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 datasets such as the Yale B or PIE datasets collected at Carnegie Mellon University. Designing methods that are robust to illumination variations in uncontrolled situations remains an open problem. Facial-expression analysis and recognition have been studied extensively in the context of human-computer interactions.[9] Facial identity and expression might be processed by separate systems. Although many techniques for automatic recognition of expressions are available, they are most effective for macroexpressions such as joy, anger, surprise, or fear.[3] Analysis and recognition of micro-expressions is an active research area that has penetrated the popular culture by way of the television show *Lie to Me!*

It's all about eyebrows

Among the different facial features, eyebrows are comparable to eyes in their importance for recognition.[1] Several possibilities arise for explaining the perceptual significance of eyebrows in face recognition.

First, eyebrows appear to be important for conveying emotions and other nonverbal signals. Since the visual system may already be biased to attend to the eyebrows

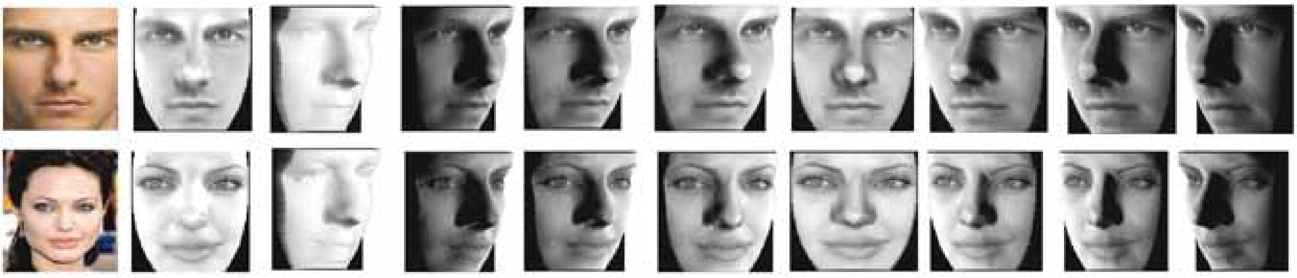

Figure 5. Examples of recovering illumination-free albedo maps and 3D models from images downloaded from the Internet. In each row, the leftmost image was downloaded from the Internet. The next two images are the reconstructed 3D models viewed from two angles. The final sets of images were generated by synthesizing new images from the 3D models that correspond to the different poses. (Image courtesy of Soma Biswas et al., *IEEE Trans. Pattern Analysis and Machine Intelligence*, May 2002, pp. 884-899.)

for detecting and interpreting such signals, this bias might also extend to the task of facial identification.

Second, for several reasons, eyebrows can serve as a stable facial feature. Because they tend to be relatively high-contrast and large facial features, eyebrows can survive substantial image degradations. Since eyebrows sit atop a convexity (the brow ridge separating the forehead and orbit), when compared to some other parts of the face, they may be less susceptible to shadow and illumination changes.

Face recognition across aging

One daunting aspect of face recognition—aging—is challenging because it must address all other variants as well. Pose, expression, and illumination changes can occur when two images of a person are taken years apart. The skin's textural properties can also differ due to makeup, spectacles, weight loss or gain, hair loss, and so on.

The facial changes that occur due to aging are influenced by several environmental factors, such as solar radiation, smoking, drug use, and stress level.[10] Biological and environmental factors can either delay or expedite the aging process. Aging causes changes in both the hard and soft facial tissues. Loss of tissue elasticity, facial volume, and alteration in skin texture also occur with aging. Although the manner of aging is highly unpredictable, a sequence of changes appears to adhere to a basic progressive pattern across time.

Drifts in facial landmarks appear to reasonably characterize the shape variations associated with aging, especially between ages 2 and 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.

Psychophysics methods include deriving cardioidal strain transformations and their extensions, variations in shape and degree of skin wrinkling, and exaggeration or a de-emphasis on facial creases. Computer-vision researchers have proposed subspace-based, model-based, and machine-learning approaches for face recognition across aging.[11] Figure 6 illustrates some appearance-prediction results derived using a craniofacial growth model.

VIDEO-BASED FACE RECOGNITION

Video-based face recognition (VFR) can establish the identity of one or several persons present in a video based on 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 a 3D model or a super-resolution image of the face, or simply learning the appearance variations from the multiple video frames.

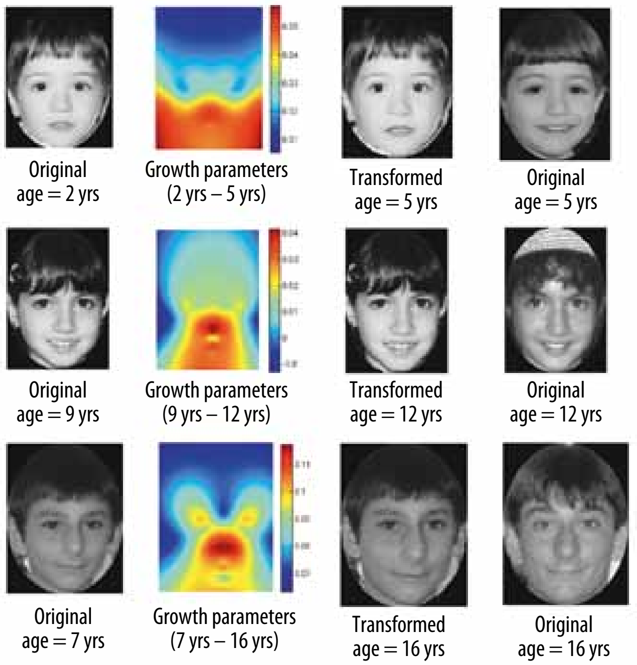

Figure 6. Some appearance prediction results derived using the craniofacial growth model.[11] The first column shows original images of children, the second estimates their growth changes, the third tracks the algorithm-estimated aging of a child, and the fourth shows an image of the child at the transformed age.

Generalizing factors

The ability to generalize across pose, illumination, expression, and other factors depends on the chosen combination. VFR is particularly useful in surveillance scenarios, in which it might not be possible to capture a single good frame, which most still-image-based methods require.

A typical VFR system acquires video feeds from one or multiple cameras, tracks and segments faces from the input feeds, extracts representations to characterize the identity of the faces in the video, then compares them with the enrolled representations of subjects in the database. This constitutes the system's test phase.

During the enrollment or training phase, a similar sequence of steps occurs, using one or multiple video feeds per identity. Researchers store the corresponding composite representations in the database. VFR approaches differ in the representation used to characterize the moving faces. An ideal VFR system performs these operations automatically, without human intervention.

Information fusion

One of the biggest challenges a VFR system faces requires effective utilization and fusion of the information—both spatial and temporal—in a video to achieve better generalization for each subject, along with

Table 2. Video-based face-recognition system examples.

Algorithm	Short description	Experimental description
Probabilistic recognition of human faces from video[x]	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[y]	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[z]	Tracks facial features defined on a grid with Gabor attributes using SIS algorithm[yy] VFR using adaptive hidden Markov models	Li dataset: 19 subjects (2 sequences each)
VFR using adaptive hidden Markov models[xx]	Statistics of training videos, and their temporal dynamics, learned by an HMM	Private: 12 subjects, Mobo[8]: 25 subjects

[x] Zhou, et al., CVIU 2003. [y] Lee et al., CVIU 2005, and Turaga et al., CVPR 2008. [z] Li et al., JOSA 2001. [xx] Liu and Chen, CVPR 2003. [yy] Li dataset: 19 subjects (2 sequences each)

discriminability across different subjects for improved identification. The fusion schemes can range from simply selecting good frames—which researchers then use for recognition in a still-image-based recognition framework—to estimation of a face's full 3D structure, which can then be used to generalize across illumination, pose, and other parameters.

The choice depends primarily on the system's operational requirements. For example, in a surveillance setting, face resolution might be inadequate for reliable shape estimation. This choice also limits the system's recognition capability. A simple frame-selection scheme cannot generalize appearance across pose variations, thus requiring the test video to have some pose overlap with the gallery videos.

Modeling facial characteristics

Effectively modeling subject-specific facial characteristics from video data can only be achieved if the changes in facial appearance during the video's course are appropriately attributed to different factors, such as pose changes and lighting and expression variations. 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, given the nature of the input data, VFR is often addressed in conjunction with tracking problems, which is itself a challenge. More often than not, tracking accuracy depends on knowledge of a reliable appearance model, which depends in turn on the identity provided by the recognition module, while the recognition result depends on the localization accuracy of the face region's input video.

Researchers have designed existing VFR systems 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 2 provides a brief summary of the existing methods.

FUTURE RESEARCH DIRECTIONS

Although there is an impressive body of experimental literature on human perception of faces, several fundamental issues are still unresolved.

Configural information

First, we must determine 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 face's overall gestalt is important. We do not, however, know exactly how to capitalize on this general notion of face gestalt. Which facial measurements contribute to this encoding?

Familiarity's role

Second, we must establish how familiarity changes facial representations. Human face-recognition processes can tolerate greater degradations in the images of people with whom we are familiar relative to those with whom we have only a casual acquaintance. This suggests that the internal facial representation undergoes important changes with increasing familiarity and thus raises several questions. What is the nature of these changes? Does encoding progress from being more piecemeal to more holistic result in a better experience? How do these changes help provide greater robustness for transformations?

Top-down expectations

Third, we must decide the role that top-down expectations on recognition will play. Recall that the response latency of a face's selective neurons in the primate inferotemporal 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? Further, under what conditions can top-down influences usefully contribute to face recognition?

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

Remote face recognition

Most existing face-recognition algorithms and systems are effective when the face images are only tens of meters from the camera. Extending the distance at which face recognition systems can be effective is a new thrust for surveillance applications. In the remote-acquisition scenario, the face images are often blurred, might not always have a sufficient number of pixels applied to the faces, and could have significant pose and illumination variations, as well as occlusion.

In the remote scenario, acquiring face signatures of sufficient quality to be fed into recognition engines is itself a challenge. This is especially true when the sensor and subjects are moving. In this case, the videos must be stabilized to robustly track the moving face before it can be recognized.

Video-based face recognition

Maritime security and other security applications require robust approaches that exploit video sequences. Video-based face recognition has received increasing attention over the past nine years. In the early stages of development, VFR research had to cope with a lack of video data for evaluation. Under the NIST Multibiometrics Grand Challenge (MBGC) program (http://face.nist.gov/mbgc), 4,489 video sequences have been made available for developing and benchmarking video-to-video matching problems.

For this work to be even more effective, the following problems must be addressed: real-time tracking and pose normalization of moving faces, illumination normalization, compensation for low-resolution face images via super-resolution techniques, and simultaneous tracking and recognition. Algorithms that can accommodate multiple gallery images or gallery and probe-video sequences must also be developed.

Face recognition in a camera network

Multicamera networks are an increasingly common solution for wide-area surveillance problems. Having a camera network acquire multiple face images can help developers build more robust descriptions of faces. This also increases the chance of the person being in a favorable frontal or near-frontal pose. However, to use this multiview information, we must estimate the pose of the subject's head. This could be done explicitly by computing the subject's actual pose 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 image resolution is low and the calibration of both external and internal cameras is insufficiently precise to allow robust multiview fusion. This is true when the subjects are usually in the far field of cameras. In addition, many other problems—such as multiview tracking, appropriate representations for multiview face images, and multiview recognition of faces—need additional investigation. Whether the algorithms must be centralized or distributed presents another major issue.

Face recognition in Web 2.0

Within the past year, face-recognition modules have been added to many applications, including Facebook, Google's Picasa, and iPhoto. These modules have been programmed to recognize faces in a person's photo library or Facebook network. They also let users correct mislabeled faces. The feedback from users will lead to rapid identification of areas in which automatic face recognition fails, revealing that we need more research.

Many of the same issues in multiview camera networks apply to face recognition in Web 2.0. However, there are unique aspects to this application, such as deriving algorithms that correctly label faces in overlapping social contact networks.

Face recognition across aging

Existing age-estimation algorithms are effective for determining ages only within a few years. The synthe-

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, if present and durable, give extremely powerful recognition cues because of the low likelihood that different people will have the exact same local attributes. Local skin irregularities have different levels of permanence. While markings such as 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 factors such as rate of blood flow to the skin and illumination variations.

Skin roughness can be divided into local and global components. Forehead and crows' feet wrinkles are examples of local surface roughness. Age or health factors may cause global roughness, which typically covers large areas of the face, such as 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.

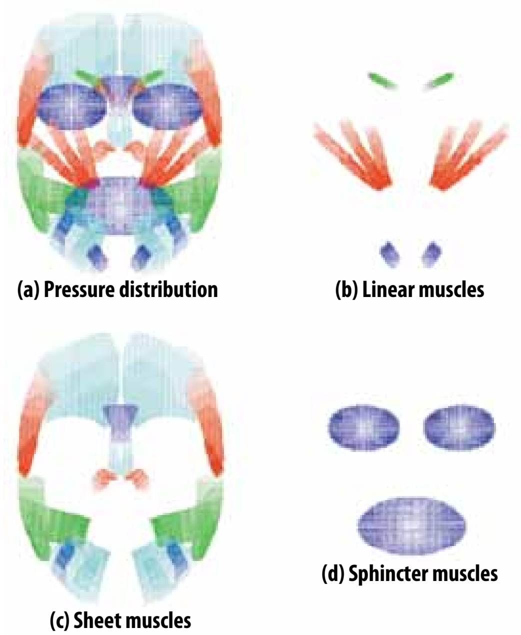

Figure 7. Illustration of facial-muscle configuration along with the accompanying proposed pressure models. The texture variation model was designed specifically to characterize facial wrinkles in predesignated facial regions.

sis of aged faces for subjects age 2 through 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 recent work,[11] developers presented brief discussions of several models for aging adults. One such model consists of shape variations and texture variations.

Attributing facial-shape variations during adulthood to the changing elastic properties of the underlying facial muscles, developers formulated the shape-variation model using physical models that characterize the functionalities of different facial muscles. Identifying facial muscles into one of three types—linear, sheet, or sphincter muscles—some authors[11] proposed transforming models for each. They modeled these facial-feature drifts as linear combinations of the drifts observed on individual facial muscles.

The texture variation model was designed specifically to characterize facial wrinkles in predesignated areas such as the forehead and nasolabial region. Figure 7 shows the facial muscle configurations and further illustrates the pressure models adopted for each type of facial muscle. The synthesis approach, however, must still be validated regarding its effectiveness for face recognition across aging.

An alternative approach that does not rely on aging models simply asks if the two age-separated faces belong to the same individual. For this approach to be useful, we must extract age-invariant features. Recent efforts have taken the nongenerative approach and aim to derive metrics for measuring the cohesiveness of feature drifts for age-separated face images of the same subject, then compare that data with age-separated faces of different subjects. Designing appropriate representations and decision rules for face recognition across aging remains an open problem.

Face and other biometrics

To ensure robustness, face-recognition algorithms have often worked in conjunction with fingerprint, iris, gait, and voice-recognition systems. This has led to the creation of a new research area: multimodal or multibiometrics systems. A salient challenge in fusing biometrics algorithms or systems is to devise efficient and robust fusion methods.

This research area has benefited largely from the theory and design of multiple classifier systems. Although researchers have advanced several examples of face/fingerprint, face/gait, face/voice, and face/iris fusion, the area of multimodal biometrics, in which the face is one of the biometric signatures, remains in its infancy.[12]

Continuing research into face recognition will provide scientists and engineers with many vital projects, including applications such as homeland security, human-computer interaction, and numerous consumer applications. The areas we are considering pursuing are recognition from unconstrained video sequences, incorporating familiarity into algorithms, modeling effects of aging, and developing biologically plausible models for human face recognition ability.

Acknowledgments

We thank Yaser Yacoob for providing the "Face Recognition: Does Skin Matter?" sidebar. Rama Chellappa acknowledges the support provided by the Office of Naval Research through MURI Grant N00014-08-1-0638. The identification of any commercial product or trade name does not imply endorsement or recommendation by the authors or their institutions.

References

1. P. Sinha et al., "Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About," *Proc. IEEE*, vol. 94, Nov. 2006, pp. 1948-1962.
2. W. Zhao et al., "Face Recognition in Still and Video Images: A Literature Surrey," *ACM Computing Surveys*, Dec. 2003, pp. 399-458.
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. M-H. Yang, D.J. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Jan. 2002, pp. 34-58.
6. P. Viola and M.J. Jones, “Robust Real-Time Face Detection,” *Int’l J. Computer Vision*, vol. 57, 2004, pp. 137-154.
7. P.J. Phillips et al., “The FERET Evaluation Methodology for Face-Recognition Algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, 2000, pp. 1090-1104.
8. V. Blanz and T. Vetter, “Face Recognition Based on Fitting a 3D Morphable Model,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Sept. 2003, pp. 1063-1074.
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*, Oxford Univ. Press, 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,” *J. Forensic Science Int’l*, Apr. 2007, pp. 1-9.
11. N. Ramanathan and R. Chellappa, “Computational Methods for Modeling Facial Aging: A Survey,” *J. Visual Languages and Computing*, May 2009, pp. 131-144.
12. A. Ross, K. Nandakumar, and A.K. Jain, *Handbook of Multibiometrics*, Springer, 2006.

Rama Chellappa *is a Minta Martin Professor of Engineering, professor of electrical and computer engineering, and an affiliate professor of computer science at the University of Maryland, College Park. His research interests include face and gait recognition, 3D modeling from videos, and activity recognition. Chellappa received a PhD in electrical engineering from Purdue University. Contact him at rama@cfar.umd.edu.*

Pawan Sinha *is an associate professor in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology. His research interests include understanding how the brain recognizes objects, scenes, and sequences. Sinha received a PhD in vision from MIT. Contact him at psinha@mit.edu.*

P. Jonathon Phillips *is an electronics engineer at the National Institute of Standards and Technology, Gaithersburg, Maryland. His research interests include face recognition, algorithm evaluation, and biometrics. Phillips received a PhD in operations research from Rutgers University. Contact him at jonathon@nist.gov.*