Reliable Face Recognition Methods

System Design, Implementation and Evaluation

Reliable Face Recognition Methods

System Design, Implementation and Evaluation

by

Harry Wechsler
George Mason University, USA

Springer

Harry Wechsler
George Mason University
Dept. of Computer Science
Fairfax, VA 22030
wechsler@gmu.edu

Library of Congress Control Number: 2006932031

Reliable Face Recognition Methods:
System Design, Implementation and Evaluation
by *Harry Wechsler*

ISBN-13: 978-0-387-22372-8
ISBN-10: 0-387-22372-X
e-ISBN-13: 978-0-387-38464-1
e-ISBN-10: 0-387-38464-2

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

With love to my children
Gabriela and Marc

Contents

Preface

Science is a way to teach how something gets to be known, what is known, to what extent things are known (for nothing is known absolutely), how to handle doubt and uncertainty, what the rules of evidence are, how to think about things so that judgments can be made, how to distinguish truth from fraud, and from show (Richard Feynman)

One of the grand challenges for computational intelligence is to understand how people process and recognize each other's face <u>and</u> to develop automated and reliable face recognition systems. This challenge underlies *biometrics*, the science of authenticating people by measuring their physical or external appearance and/or their behavioral or physiological traits. The physical and behavioral traits are not necessarily independent. The face we look at is a mix of both physical characteristics and emotive expressions. Face recognition has become a major biometric technology. Solving the face recognition problem will have a major scientific impact, as recognizing people is a first but critical step towards building intelligent machines that can function in human environments. "The ability to recognize living creatures in photographs or video clips is a critical enabling technology for a wide range of applications including defense, health care, human-computer interaction, image retrieval and data mining, industrial and personal robotics, surveillance and security, and transportation. Despite 40 years of research, however, today's recognition systems are still largely unable to handle the extraordinary wide range of appearances assumed by common objects [including faces] in typical images" (*Designing Tomorrow's Category - Level 3D Object Recognition Systems* [1]).

Biometrics has become the major component in the complex decision-making process associated with security applications. Key concerns related to accuracy and performance, benefits versus costs, information assurance, and security over privacy have surfaced and have yet to be resolved. Skepticism, the heart of scientific method, is needed to ferret out fact from fiction regarding what biometrics can actually do and to what extent. Advancing the field of biometrics for homeland security has taken on a sense of urgency in the post 9/11 world. Even though people can detect and identify faces with little or no effort, building an automated system for such purposes has proven elusive as reliable solutions have yet to be found. The all-encompassing *Face in a Crowd* biometric problem addresses both face detection and face recognition in cluttered environments. Biometric systems have to take into account the dynamic changes that affect the visual stimuli, including variability in the geometry of image formation, such as facial pose and distance from the camera, and illumination. Other factors that affect face recognition include facial expression due to emotion, occlusion and disguise, temporal changes and aging, and last but not least, the lack of adequate training data for learning how to represent and encode human faces.

A few major edited books treat face recognition. Among them are the first and seminal "*Face Recognition: From Theory to Applications*" (Wechsler et al., Springer, 1998), and most recently the "*Handbook of Face Recognition*" (Li and Jain, Springer, 2005). This book is the first to comprehensively address the face recognition problem in its entirety, while drawing inspiration and gaining new insights from complementary fields of endeavor, such as neurosciences, statistics, signal and image processing, computer vision, machine learning and pattern recognition, and statistical learning. The various chapters treat topics related to how people represent, process and/or respond to the human face, modeling and prediction, the face space, identification and verification, face detection, tracking and recognition, 3D, data fusion, denial and de-

[1] http://lear.inrialpes.fr/people/schmid/workshop.html

ception using occlusion and disguise, performance evaluation and error analysis, and finally, competing security and privacy considerations.

The underlying theme of the book is that the biometric inputs chart continuous and coherent space and time manifolds, which facilitate their recognition. Face recognition is dynamic rather than static. It continuously iterates, making specific interpretations and assigning confidence to them. Supporting and non-accidental evidence is accrued in an active fashion, leading to lower uncertainty in the recognition decisions made, and resolving ambiguity, if any. Integral to face recognition are advances in pattern recognition. Novel methods are proposed here to handle real life applications where variability, incomplete, noisy, distorted and/or disguised patterns are usually the norm rather than the exception. The overall goal of the book is *applied modern pattern recognition*, with the understanding that the novel methods described here apply to any objects. The face pattern is only one of the object patterns that surround us and need to be recognized. The scope for pattern recognition (Rosenfeld and Wechsler, 2000) is much wider here because among other things both training and testing can take place using incomplete or camouflaged/disguised patterns drawn from single and/or multiple image sets.

The emphasis throughout the book is on proper modeling and prediction. Gregory Chaitin, in the March 2006 issue of the Scientific American, recalls Gottfried Leibniz's 1685 philosophical essay *Discourse de métaphysique* (Discourse on Metaphysics). The essay discusses how one can distinguish between facts that can be described by some law and those that are lawless, irregular facts. Leibniz observed that "a theory has to be simpler than the data it explains, otherwise it does not explain anything. The concept of a law becomes vacuous if arbitrarily high mathematical complexity is permitted, because then one can always construct a law no matter how random and patternless the data really are." The corollary for Chaitin is that "a useful theory is a compression of the data; comprehension is compression." Modeling and prediction, the hallmarks of learning, can be implemented using novel methods driven by semi-supervised learning and transduction, exchangeability and rank order, and martingale. Overall, the book articulates new but promising directions for pattern recognition, while providing the motivation for innovative ways to approach the face recognition challenge.

The title of the book, *Reliable Face Recognition Methods*, refers to the normal expectation one has that face recognition should display robust performance despite suboptimal and/or adverse image acquisition conditions or lack of adequate training data. Even the top-ranked face recognition engines still reject legitimate subjects, while letting impostors pass through. "Reliable," throughout the book, means the ability to deploy consistent, dependable, large-scale and full-fledged operational biometric systems, which is the true hallmark of a mature technology. To that end, a large data base of facial images, such as FERET, is required to test and assess competing technologies. FERET, which was designed and developed at George Mason University under my direction, has become the standard data base used by researchers worldwide for R&D and benchmark studies on face recognition. Science needs to be replicated and tested for validation.

This book can serve both as an interdisciplinary text and as a research reference. Each chapter provides the background and impetus for understanding the problems discussed and the approach taken to solve them. The book can benefit advanced undergraduates ("senior") and graduates taking courses on pattern recognition or biometrics; scientists and practitioners interested in updating their knowledge; and government and industry executives charged with addressing ever-evolving biometric security requirements.

My gratitude goes to many people. Many thanks go to Professor Jack Sklansky, who introduced me to the field of pattern recognition. Much appreciation goes to my former doctoral students Srinivas Gutta, Shen-Shyang Ho, Jeffrey Huang, Fayin Li, and Albert Pujol, with whom I had many productive and rewarding collaborations. I am also grateful for the help and inspiration for this book from Josef Bigun, Vladimir Cherkassky, Clifford Claussen, Victor Chen, Stephen McKenna, Matt Matsuda, and

Barnabas Takacs. My thanks go also to the many people referenced throughout the book from whom I have drawn knowledge and motivation. From my brother Tobi, I learned to appreciate and love the visual arts, which led me to explore the scientific basis for perception-representation and interpretation. Thanks go also to my sister-in-law Nobuko for her friendship and kindness. Last but not least, heartfelt thanks go to my wife Michele for her encouragement and help with completing this book, and to my children Gabriela and Marc for the sparks in their eyes and their smiling faces.

Harry Wechsler
June 2006

1. Introduction

The first rule was never to accept anything as true unless I recognized it to be certainly and evidently such: that is, carefully to avoid all precipitation and prejudgment, and to include nothing in my conclusions unless it presented itself so clearly and distinctly to my mind that there was no reason or occasion to doubt it. The second was to divide each of the difficulties which I encountered into as many parts as possible, and as might be required for an easier solution. The third was to think in an orderly fashion when concerned with the search for truth, beginning with the things which were simplest and easiest to understand, and gradually and by degrees reaching toward more complex knowledge, even treating, as though ordered, materials which were not necessarily so. The last was, both in the process of searching and in reviewing when in difficulty, always to make enumerations so complete and reviews so general, that I would be certain that nothing was omitted.

From Discourse on Method and Meditations by Ren Descartes (1641) (translated by Laurence J. Lafleur and published by Liberal Arts Press, 1960)

Face recognition (Samal and Iyengar, 1992; Chellappa et al., 1995; Daugman, 1997; Jain et al., 1999; Zhao et al., 2003; Bolle et al., 2004; Li and Jain, 2005) has become a major biometric technology. Biometrics involve the automated identification or authentication from personal physical appearance or behavioral traits. Human physical appearance and/or behavioral characteristics are counted as biometrics as long as they satisfy requirements that include universality, distinctiveness or uniqueness, permanence or invariance, collectability, and acceptability (Clarke, 1994). The early belief in the uniqueness aspect of faces (to preempt forgeries) was one of the reasons behind their use, e.g., the face of Queen Victoria on the early stamps (Samal and Iyengar, 1992). Biometric systems, including face recognition systems, can be categorized according to their intended applications. According to Wayman (1999) a suitable self-evident taxonomy will include cooperative vs. non-cooperative, overt vs. covert, habituated vs. non-habituated, attended vs. non-attended, standard vs. non-standard operating environments, and public vs. private.

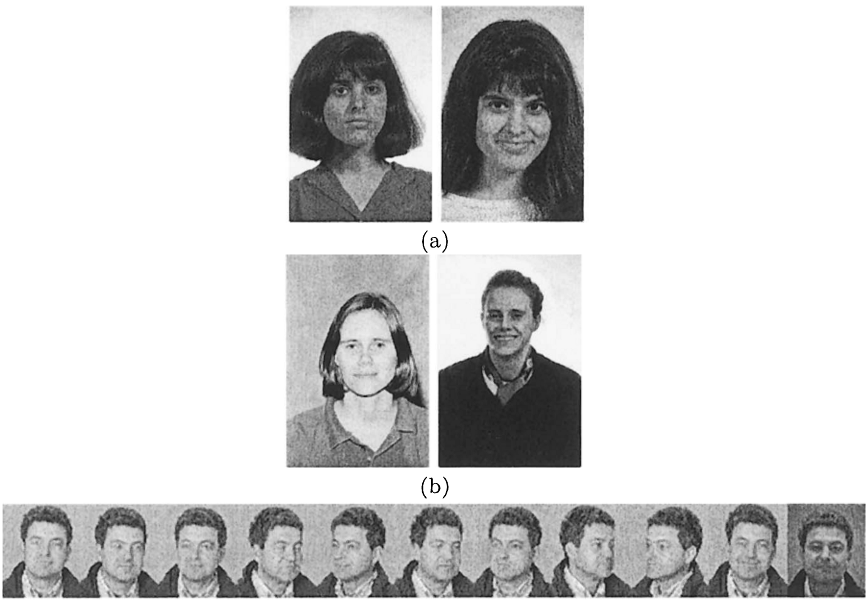
This book addresses the above taxonomy as it discusses face recognition along the complementary dimensions of science, (information) technology and engineering, culture and society, and visual arts. It is about science because it aims to understand and systematize the fundamental principles behind face processing. Face processing is an all-encompassing term that involves everything that facilitates face recognition, e.g., image capture, enrollment, and face detection and tracking. The scientific dimension is related to the basic research that supports technological progress. The book is about technology and engineering, because it deals with applied science and research aimed at practical ends, e.g., designing and building reliable face recognition systems. It is about culture and society because they affect the role the human face plays in our interactions. The book is also about the visual arts because the human figure has always occupied an important place in personal expression and contemplation. Art connects internal and external realities, provides for new perspectives of the world, and seeks for the ultimate truth and permanent essence embodied by fixed icons such as human faces and ideals. The arts activate abstraction and creativity and can stimulate innovative face recognition research, e.g., using the golden ratio template of human beauty for face recognition-by-parts (See Sect. 6.5 and 9.7). Last but not least, the book is about building completely automatic and full-fledged biometric systems that consider the full range of the face recognition sub-tasks, starting with data acquisition and enrollment and ending with different face authentication scenarios. The ever expanding scope of face recognition includes field operation rather than controlled in vitro lab conditions. This should lead to building search and analysis biometric video engines able to recognize people and/or interpret their activities and intentions from live-fed CCTV.

The book, multidisciplinary and syncretic, frames a modern research agenda for pattern recognition, in general, and face recognition, in particular. The modernity aspect refers to the scope of the enterprise. The book identifies real problems and motivates the need for large scale pattern recognition systems that can handle human diversity, temporal changes, and occlusion and disguise. The book, selective rather than encyclopedic, introduces new topics that require further investigation. It differentiates and motivates among existing problems and their proffered solutions, places emphasis on common threads, and focuses on what is most important. In particular, the book aims to fuse and reconcile the specific disciplines of image and signal processing, computer vision, machine learning and pattern recognition, while charting new but promising research directions.

Computer vision is about "computing properties of the 3D world from one or more digital images. As the name suggests, it involves computers interpreting images. Image analysis and/or understanding are synonyms for computer vision. Image processing and pattern recognition are disciplines related but not identical to computer vision. Image processing concerns image properties and image-to-image transformations, while pattern recognition [involves] recognizing and classifying objects using digital images" (Trucco and Verri, 1998). Learning, which is about generalization and prediction, lies at the interface between computer vision and pattern recognition. It plays a fundamental role in facilitating "the balance between internal representations and external regularities" (Nayar and Poggio, 1996). Face recognition requires new and robust learning paradigms. This includes 'good' classification methods that can work with only limited training data, which was acquired under fairly flexible and general assumptions. "The fewer assumptions a [computer vision] system imposes on its operational conditions, the more robust it is considered to be" (Moeslund and Granum, 2001).

The challenges confronting face recognition are many. First and foremost there is much variability in the image formation process that includes geometry, illumination, occlusion and disguise, and temporal changes (see Fig. 1.1). Even the faces of "identical" twins are different to some extent (see 1.1a). Biometrics in general, and face recognition, in particular, bear directly on the use of forensics in the courts of law. In a provocative *Science* editorial, titled "*Forensic Science: Oxymoron?*" Donald Kennedy, the Editor-in-Chief, makes the obvious point that the reliability of forensics "is unverified either by statistical models of [biometric] variation or by consistent data on error rates. Nor does the problem with forensic methods ends there. Processing and enhancement of such images could mislead jurors who believe they are seeing undoctored originals." Following the 1993 U.S. Supreme Court's Daubert case, the Court "did list several criteria for qualifying expert testimony: peer review, error rate, adequate testing, regular standards and techniques, and general acceptance" (Kennedy, 2003). Similar arguments apply to automatic face recognition and are considered throughout. Other factors adversely affecting face recognition include the lack of enough data to learn reliable and distinguishable face representations, and the large computational resources required to adequately process the biometric data. Comprehensive evaluations of the science underlying forensic techniques in general, and studies on the uniqueness of personal face signatures, in particular, are still lacking. The current Face Recognition Grand Challenge (FRGC) project (Phillips et al., 2005), administered by the US National Institute of Standards and Technology (NIST), aims for

98% average reliability at FAR = 0.1%, "a tough standard, but perhaps not tough enough to handle tens of millions of travelers per year", when one considers the false alarms. The scope for FRGC is relatively narrow compared to the earlier FERET and FRVT evaluations (see Sect. 13.4) because despite the relatively large corpus of data involved, the number of subjects enrolled, 275 and 675 for ver1.0a and ver2.0, respectively, is only in the hundreds and thus much smaller than FRVT2002. FRGC functionality is further limited to verification compared to previous evaluations that also involved identification. Last but not least, the practicality of FRVT during both enrollment and testing is questionable due to its requirement for a large set of face images using different image capture methods.



(a)

(b)

(c)

Fig. 1.1. Human Faces (from FERET). (a) Twins; (b) Temporal Variation; (c) Time Sequence Including Pose Variation.

The book is no panegyrics to some research utopia but rather an attempt to be as informative as possible, avoid heuristics, and last but not least cope with meaningful face recognition tasks (see Descartes' admonishments). The book is inclusive but in a comparative fashion in order to motivate and inspire. Hard or intractable problems, e.g., correspondence, segmentation (Gurari and Wechsler, 1982) and clutter, variability, and/or insufficient and/or missing information, are not avoided or glossed over. Folk wisdom chuckles that the difference between theory and practice finds no difference in theory but only in practice. Vapnik (2000) rightly points out, however, that there is nothing more practical than a good theory. The challenge for reliable face recognition is to show through fair and good experimentation that theory and practice are consistent.

The book is mostly about face recognition but it is quite relevant to categorization and recognition for science and technology in general. The driving and unifying force behind the proposed reconciliation among computer vision, machine learning, and pattern recognition, is the active, progressive and selective accrual of evidence needed to reduce uncertainty. Practical intelligence "modifies the stimulus situation as a part of the process of responding to it" (Vygotsky, 1976). Practical intelligence is actually much more than merely modifying or transforming the input. "For the young child, to think means to recall; but for the adolescent, to recall means to think" (Vygotsky, 1976). Connect "adolescent" to reliable face recognition engines, and connect "think" to reasoning and inference. Faces cannot be reliably located and identified from merely one still image. Rather than static inputs, the language of changes observed, their logical interrelations, and the plausible inferences or transformations over space and time underlie reliable face identification and authentication.

1.1 Tasks and Protocols

The generic (on-line) biometric system used herein for face recognition (see Fig. 1.2) will be referred to throughout the book. The **match** task evaluates to what extent the biometric **signatures** extracted from the unknown face exemplar(s) and the biometric signature(s) stored during the enrollment stage as reference **template(s)** are similar. The match score has to be compared against some a priori defined **threshold** value. Matching takes place against a single template (for **verification**), or against a list of candidate templates (for **identification**). Verification is also referred to as **authentication**. Identification is usually carried out using iterative verification and ranking. The face space, usually the basis needed to generate the templates, is derived using face images acquired ahead of time and independent of those that would be later on enrolled for training or queried on (see top of Fig. 1.2). The biometric templates encode the essential features of the face along the specific dimensions of the face space used. They are stored in some central data base but can be also carried by owners on a smart card.

Face recognition performance is still lacking. According to the December 6, 2003 issue of the Economist "governments are investing a lot of faith in biometric technology as a way to improve security. For the moment, this confidence is misplaced. Body-recognition technology is not the silver bullet many governments imagine it to be. Biometrics [are] too flaky to trust." The experience of the 2001 Super Bowl held in Tampa and the trial held at Boston's Logan International Airport in 2002, which exhibited a failure rate of 38.6% [while the false-positive rate exceeded 50%], are cases in point. Again according to the Economist "given the volume of air traffic, the incidence of false alarms will vastly outnumber the rare occasions when someone tries to subvert the system. The false alarms will either have to be ignored, rendering the system useless, or a time-consuming and expensive secondary-screening system will be needed." This book is about how to improve the state-of-the art for reliable face recognition.

Performance evaluation is an integral part of any serious effort to field reliable face recognition systems. **FERET** (Phillips et al., 1998) and **BANCA** (Bailly-Bailliere et al., 2003), the standard evaluation protocols in use today, are briefly described next. FERET starts by considering target (**gallery**) T and query (**probe**) Q sets. The output for FERET is a full (distance) matrix $S(q,t)$, which measures the **similarity** between each query face, $q \in Q$, and each target face, $t \in T$, pair. The nearest neighbor (NN) classifier authenticates then face images using the similarity scores recorded by S. The availability of the matrix S allows for different "virtual" experiments to be conducted when one selects the specific query P and gallery G as subsets of Q and T. Note that one can expand on the above model using data fusion when sets rather than singletons are matched, and both the query and the gallery sets are acquired using multimodal sensors.

The **closed (universe) set face recognition** model used by FERET for 1 : N identification, when each probe has always a mate in the gallery, is restrictive and does not reflect the true intricacies of positive and negative biometric enrollment and identification. Under positive enrollment, the client is authenticated to become eligible for "admission" or apprehended if found on some watch list, while under negative identification the biometric system has to determine that the client does not belong to some most-wanted list. Positive identification can be determined using traditional personal tokens, e.g., PIN, but negative identification can only be established using biometrics.

More challenging is the **open (universe) set face recognition** model, which operates under the assumption that not all the probes (unknown test face images) have mates (counterparts) in the gallery (of known subjects) (see Sect. 6.3). Open set face recognition requires the **a priori** availability of a **reject** option to provide for the answer "none of the above" for unknown classes of clients. If the probe is detected rather than rejected, the face recognition engine must then identify/recognize the subject. The operational analogue for open set face recognition is the (usually small) **watch list** or **surveillance** task, which involves (i) negative identification ("rejection") due to the obvious fact that the large majority [almost all] of the people screened at security entry points are law abiding people, and (ii) correct identification for those that make up the watch list. "Performance for the open set problem is quantified over two populations. First the impostors, those persons who are not present in the gallery, i.e., not on the watch list, are used to compute the false match [acceptance] rate, which is needed to quantify rejection capability. Second, for those persons who are "known" (i.e., previously enrolled) to a system, the open set identification rate, is used to quantify user [hit] performance" (Grother, 2004).

The 1 : N open set problem referred to by FRVT2002 (Phillips et al., 2003) as the watch list task, is briefly addressed after two (degenerate) special cases of verification and closed set identification. Verification corresponds to an open set identification for a gallery size of $N = 1$, while closed set identification seeks the match for an image whose mate is known to be in the gallery, i.e., for each image probe $p \in P$ there exists (exactly one) gallery mate $g^* \in G$. The "none of the above" answer is not an option. Cumulative Matching Curves (CMC) and Receiver Operating Characteristics (ROC) are used to display the results for identification and verification, respectively (see Sect. 12.1). FERET results are derived using ground truth for a posteriori setting of optimal thresholds to yield prespecified false alarm rates. Ground truth, however, is not available during real field operation hence the need for a priori setting of decision thresholds.

The BANCA protocol, geared toward the verification task, is designed to work with multi-modal databases. Verification is viewed as hypothesis testing and the (detection) choice is between true clients and impostors. There are two types of errors, false acceptance and false rejection, and their associated costs. Two types of protocols exist, closed and open set, respectively. In closed set verification the population of clients is fixed and anyone not in the training set is considered an impostor. The earlier XM2VTS Lausanne protocol (Bengio et al., 2001) is an example of closed set verification. In open set verification one seeks to add clients without having to redesign the verification system. In particular, BANCA goal is to use the same feature space and the same design parameters including thresholds. In such a scenario, the feature space and the verification system parameters should be trained using calibration data distinct and independent from the data used for specifying the client models (see Fig. 1.2). The BANCA protocol is an example of open set verification protocol.

The use of the open set concept by the BANCA protocol is quite restricted. It only refers to the derivation of the feature (face) space and the parameters needed for verification. This was referred earlier as face space basis derivation (see top of Fig. 1.2 and Sect. 5.1) and should precede enrollment. BANCA protocol, however, does not address the full scope of open set identification, where not all the probes are mated in the gallery. Real world applications are of the open set type. We address this important but usually neglected aspect of face identification using transduction, a local form of estimation and inductive inference (see Sects. 6.3 and 6.4).

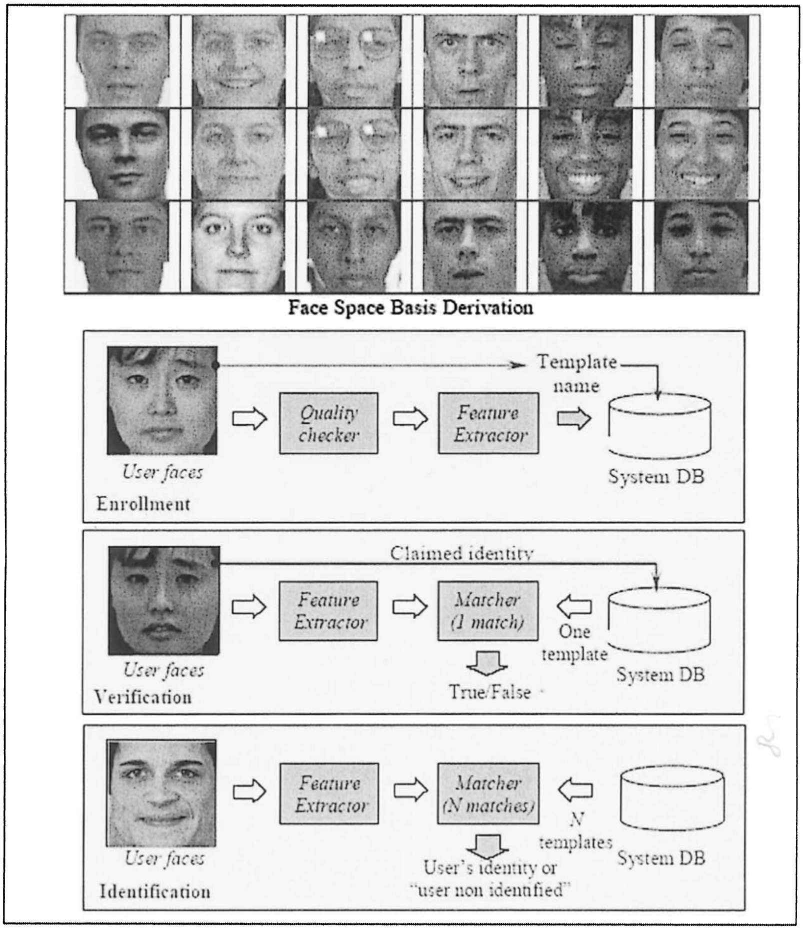


Fig. 1.2. Face Recognition Methodology (Reprinted from Li and Wechsler, Open Set Face Recognition Using Transduction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ©2005 IEEE).

1.2 Biometrics System Design

The biometric system is the technological shell needed to realize the face recognition methodology. The focus for biometric technologies is much more than simply "replacing passwords" to being "fundamental components of secure systems, whose use and integration demands very careful planning. This involves the consideration of many issues, such as recognition accuracy, total cost of ownership, acquisition and processing speed, intrinsic and system security, [exception handling], privacy and legal requirements, as well as interface usability and user acceptance" (Bolle et al. 2004).

There are many constraints, both logistical and technical, the biometric system needs to heed. From a logistical view point most important is to become sensitive to the facts that (i) biometrics serve mass screening, e.g., machine readable travel documents (MRTD), and that (ii) the number of wanted and/or impostors is relatively small compared to the large majority of law abiding citizens. The road map for the biometric system is drawn by system engineering (Kotonya and Sommerville, 1998). It involves a complex life cycle process that includes requirements engineering, system prototyping and evaluation, architectural design and specifications, subsystem development (biometrics, communication, storage and retrieval, and security and privacy), system integration, system validation, and system maintenance. The requirements engineering component involves process together with design and techniques, i.e., *what* has to be done and *how*. System engineering addresses the requirements engineering aspect using a *spiral process* that consists of requirement elicitation, analysis and negotiation, requirements documentation and validation, and requirements management (documentation and maintenance). The design criteria used include interoperability and standards, policies, vulnerability and security, privacy, costs/performance/benefits tradeoffs, user acceptance and human factors, documentation and support, software development kit (SDK), and system administration and maintenance.

Biometric systems can be viewed as pattern recognition systems. The specific biometric components include data collection, enrollment and storage, feature extraction and template derivation, matching for different face recognition tasks such as identification and/or verification, and decision-making including post-processing. The matching component is usually referred to as the face recognition engine. Enrollment assumes that the engine has been trained ahead of time and taught how to generate the template signatures needed for enrollment and later on for matching and recognition. The biometric signatures have to be securely stored and transmitted and their vulnerabilities carefully assessed (see Ch. 14).

System validation involves testing and evaluation "to help testers achieve the best possible estimate of field performance while expending the minimum effort in conducting their evaluation, and to improve understanding of the limits of applicability of test results and test methods" (Mansfield and Wayman, 2002) (see Chaps. 12 and 13). Blackburn (2001) provides a structured approach that moves system validation through three major steps: a **technology evaluation**, a **scenario evaluation** and an **operational evaluation**. "Each type of [evaluation] test requires a different protocol and produces different results. [Technology refers to the algorithms used.] The goal for scenario evaluation is to determine the overall system performance in a prototype or simulated application. Testing is carried out [on-line] and using a complete system in an environment that models a real-world target application of interest. In scenario and operational testing any adjustments to the devices and their environment for optimal performance (including image quality and decision thresholds) have to take place **prior** [emphasis added] to data collection [and enrollment]. This should be done in consultation with the vendor" (Mansfield et al., 2001). Alternatively, according to Wayman, technology, scenario, and operational settings correspond to testing the algorithms, testing the human-machine interface, and testing mob behavior, respectively. This requires off-line training while enforcing specific policy management rules for field deployment and operation. Note that FRVT2002 has undertaken only (algorithmic) technology and scenario evaluations.

Deployment of operational biometrics involves an a priori but necessary step of threshold selection. This step is difficult to automate due to the strong dependency of optimal thresholds on image quality and the composition of training data. Much more is also known about the population, [or genuine customers,] of an application than about the enemies, [i.e., the impostors]. Consequently, the probability of a false alarm rate (FAR), [i.e.,] a false match [for screening and positive identification], is hard to estimate. Hence, "the false reject rate (FRR) for a particular decision is easier to estimate than the false alarm rate for that decision, because the biometric samples of the enemy population are not available" (Bolle et al., 2004). Note that FAR and FRR are also referred to as FPR (false positive rate) and FNR (false negative rate), respectively.

The Biometric Evaluation Methodology (BEM) provides "a common criteria scheme for Information Technology Security Evaluation that is mutually recognized by organizations in several countries" [1]. The administrator guidance [AGD] provided by the BEM document specifically refers to the setting of decisions thresholds and notes that "where it is possible to change the matching thresholds used in the comparison [authentication] process, documentation should include the effects of changing these thresholds, the means of changing these thresholds, and the importance of these thresholds in determining security." The decision threshold must be considered to be a security parameter according to BEM (AGD_ADM. 1-5). It must also include, according to BEM - Vulnerability Assessment and Misuse (AVA_MSU2-10) "guidance on how to set matching thresholds, if changing the threshold is permitted." Biometrics systems in general, and face recognition engines, in particular, require significant training for tuning and calibration before "plug and play" becomes feasible, if at all.

1.3 History

An NSF sponsored international workshop, held late in 2003, on *Designing Tomorrow's Category-Level 3D Object Recognition Systems* had a rather grim assessment of the state-of-the art for object recognition. While a distinction was made between identification (recognizing the same object), e.g., personal authentication, and categorization (recognizing a visual object class), e.g., gender and/or ethnicity, no mention is made of personal identification that discriminates between instances from the same category, e.g., human faces that carry different ID tags. The NSF report claims that "for humans, categorization is easier whereas in computer vision individual identification is a much simpler problem. For example, humans can recognize categories such as dogs, cats, horses etc. by the age of three. It appears that in humans categorization precedes [generic rather than personal] identification." The reality for face recognition is quite different. The performance on biometric categorization far exceeds that for personal identification, because there are less biometric categories, and the within class variability, characteristic of human identity, is far greater than that of biometric categories.

With the advent of photography by mid-19th century, police departments started to build local archives of suspected felons. At the beginning, the gallery of sought after criminals included daguerreotypes or mug shots together with associated information, something that today is referred to as soft biometrics. Galleries were compiled by both local police and by private detective services, such as the Pinkerton National Detective Agency. Alphonse Bertillon (1853–1914) is widely acknowledged as the person who started the field of biometrics. A friend of Paul Broca, he held the firm belief that the physical or anthropometrical characteristics of people, e.g., body measurements, are unique and that they are measurable. Specious links between anthropology and criminology, e.g., phrenology [shape of the skull and face reveals mental character and capacity], did emerge too. Bertillon measured various features, including height, arms' length, and the length and breadth of the skull. Beyond obvious bodily measurements,

[1] http://www.cesg.gov.uk/site/ast/biometrics/media/BEM_10.pdf

Bertillon's system also included morphological descriptions of body parts including the ear, and marks due to disease, accident, or personally inflicted, e.g., tattoos. "Having made his first 7,336 measurements, Bertillon [handwriting expert at the Dreyfus trial] was able [in 1883] to identify 49 repeat offenders; the following year, 241" (Matsuda, 1996). Bertillon intended to use these biometrics for the Paris prison service, for both physical and "moral" identification, and in 1892 became the first Director of the Paris' *Service d'Identite Judiciare*. Taxonomies across demographics and physical space were also produced. The core of *fiche signaletique* designed by Bertillon, similar to recognition-by-parts today, was to decompose the face first into "its [character] traits" before recomposing it later on for final identification. A culturally bias-based approach is a core weakness of his approach. "Cognition dominated perception, and the eye searched for what it expected to find, that is *resemblance*. Seeing a face was a fine thing, but such memories were easily tricked. According to Bertillon *we only think that which we are able to express in words*" (Matsuda, 1996) (see Sect. 14.6 on photofits).

Long before the Daubert case facing the US Supreme Court, Bertillon was aware of both the difference between early enrollment and delayed identification and authentication, and of the importance that uniqueness plays in biometrics. Towards that end, Bertillon calculated that if 14 different measurements of body parts were taken, the odds of finding two people with identical measurements were 286,435,456 to one. The system proposed by Bertillon did not last for long. In 1898, a Royal Commission sided with Sir Francis Galton - responsible for establishing fingerprints as the method of choice for personal identification and authentication - who argued that the statistical interpretation used by Bertillon is flawed and "the incorrectness lay in treating the measures of different dimensions of the same person as if they were independent variables, which they are not. For example, a tall man is much more likely to have a long arm, foot, or finger, than a short one." In addition to being conceptually flawed, Bertillon's approach failed because it was unreliable and difficult to administer.

As early as 1888, Galton (1888a and 1888b; 1910) proposed a method to classify faces, by indexing different facial (curve) profiles, finding their norms and bounding limits, and classifying new faces from their deviations from the norm. Five cardinal points derived from the face profile were used for matching. The points used included "the notch between the brow and the nose, the tip of the nose, the notch between the nose and the upper lip, parting of the lips and the tip of the chin" (Samal and Iyengar, 1992).

Anthropometry, which includes standardized body measurements, archival and retrieval, came to the United States courtesy of Major McClaughry in 1887. Gates (2004), drawing from Matsuda (1996), further reconstructs some of "the past perfect promise of facial recognition technology." In 1907, medical-legalist Dr. Motet in France transformed "bodies into coded numerical references and circulated those references as a system of signals via telegraph," while in 1912, Dr. Icard, also in France, proposed "bodies as figural landscapes to be plotted cartographically." Soon after Roentgen discovered X-rays, Dr. Foveau de Courmelles suggested the "internal card" as the ultimate identity trace that is least changeable, and established the link between biometrics and personal forensics.

Personal documentation came into vogue around the turn of the 20th century, a time of great mass dislocation when communal identity was shifting fast. The appeal for identity cards and passports, standardization and search efficiency, applied to vagabonds and foreigners alike. Matsuda (1996) recounts Vacher's sensational criminal case, which at the turn of the century ended on the guillotine in 1897. What finally trapped Vacher was "the memory of the state," which Max Weber called "the domination through knowledge, specifically rational" that characterizes modern bureaucratic organizations, i.e., the knowledge of the "file." The "file" includes photographs, measurements, [identity cards] documents, clues and correlations, ultimately "collapsing the distinction between *identity* and *identification*. The memory did more than remember - it was a *memory* which acted." The implications today, for face recognition on one side, and security and privacy, on the other side, are clear. Multi-modal face recognition, soft biometrics, evidence accumulation, and data fusion and integration further the identification and authentication process but at the cost of decreased privacy.

An excellent review for the first 25 years of the modern age, including the contributions made by Galton, comes from Samal and Iyengar (1992). The modern age for face recognition started with Bledsoe (1964). The approach used by Bledsoe was the (local) *feature* or *abstractive approach.* The intuition suggested looking at the major features of the face and comparing them against the same features from the candidate faces available. Fiducial landmarks were located and extracted as major features using manual annotation, e.g., the corners of the eyes, ears, nose, and the mouth. Normalized distances and ratios, computed from these fiducial landmarks to a common reference point, were compared to labeled data for classification. Preston (1965) used the matched filter and correlation (see Sect. 10.6) for identification from a set of six kings. In the late 1960s, Ilya Prokopoff, a Soviet scientist from Moscow University, approached face recognition using hybrid methods that included models of neurons and Rosenblatt's Perceptron. Faces were scanned and matched against a visual database to find local correlations. To address the uniqueness problem Kaya and Kobayashi (1972) used information theory to determine the number of faces that Bledsoe's method could distinguish.

Sakai et al. (1969) were first to program the computer to locate faces in an image. The feature extraction approach continued with Kelly (1970), whose doctoral dissertation was the first to show how to automatically extract outlines for both the head and body, and how to proceed to locate the facial landmarks. Body measurements, e.g., width of shoulders, and close-up head measurements, e.g., head width and the interocular distance, were combined to identify about ten persons. Leon Harmon (1973) asked whether a computer can be programmed to recognize a human face. Faced with the computational limitations of the time, he experimented with "block portraits" and asked what is the minimum resolution required for effective face recognition. He experimented with blurring - progressive defocusing - of photos into coarse 16×16 images using low-pass frequency filters (see Sects. 2.2 and 5.2) and displayed with 8 or 16 gray levels. This process can result in the image shown below (Fig. 1.3) where one can recognize Abraham Lincoln. Another role frequency plays is seen in critical-band masking when adjacent noise lies within two octaves from the signal. Harmon and Julesz (1973) showed that critical-masking is responsible for suppressed recognition (see also Sect. 14.5 on anonymity). The above findings supplemented by additional observations (see Sect. 2.2) suggest that frequency analysis plays a major role in face recognition.

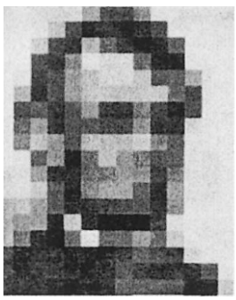


Fig. 1.3. Blurred Image of Abraham Lincoln.

The next question Harmon asked was about what features are most important for face recognition. The sketches drawn by police to identify missing or wanted people start from a catalogue of face portraits drawn from memory and indexed by various head shapes, eye spacing, and lip thickness (see Sect. 14.6 for photofits). The sketching process, which iterates using "pointing to" similar features on other portraits, yields a written description used to create the final sketch. Significant facial details are corrected to enhance discrimination. Enhancing and/or distorting some of the facial features is yet another possibility characteristic of caricatures (see Sect. 5.6). While blurring serves to cancel noise, e.g., sharp edges due to digitization, high-pass filtering sharpens the visual appearance of the features that are deemed most important.

Harmon also built a vocabulary of features and the values they can take for describing and ultimately recognizing faces. The features expanded on the fiducial landmarks used almost one hundred years earlier by Galton and more recently by Bledsoe. They included hair (coverage, length, texture, and shade), forehead, eyebrows (weight and separation), eyes (opening, separation, and shade), ears (length and protrusion), cheeks, nose (length, tip, and profile), mouth, lip (upper - and lower thickness, overlap, and width), and chin (profile). Examples of values the features could take include straight, wavy and curly for hair texture, and sunken, average, and full for the cheeks. Matching between two faces, or equivalently between their feature vectors, was done using the Euclidean distance normalized for the variance observed in the placement of features. The profiles were drawn by artists but the fiducial landmarks were automatically derived (Samal and Iyengar, 1992). Kanade (1973) was the first to program the computer to recognize faces using the feature based approach. He used dedicated subroutines for different parts of the face and was able to automatically identify 15 out of 20 people. Kaufman and Breeding (1976) used profiles for identification. The approach was feature based and the features were extracted from the autocorrelation function expressed in polar form.

The next approaches for face recognition were *holistic* or *template-matching* and *hybrid*, respectively. The holistic approach, *global* in nature, encodes the whole face and achieves identification using template matching or correlation (Baron, 1981). There is unity to the face and the Gestalt or whole is more than the sum of its components. Faces are perceived as a whole rather than disconnected features. The reason for the holistic approach came from the realization that seeking for more and better features is not feasible. More measurements are difficult to come by in an automatic fashion and their quality deteriorates due to noise and occlusion (Brunelli and Poggio, 1993). Eyes were located first using templates via correlation. Standard normalization for faces kept the inter (between the eyes) ocular distance constant for storage and later on for retrieval using correlation. The best candidates were found using template matching. Disadvantages for the global approach include the need for extensive training and the difficulty to interpolate between exemplars and models, e.g., poses. The top ranked faces for global methods can continue to compete using feature matching. The last approach is referred to as *hybrid* due to its use of both template and feature matching.

The field of face recognition was reinvented when Kirby and Sirovich (1990) proposed Principal Component Analysis (PCA) for holistic face recognition. PCA, conceptually similar to Karhunen-Loeve (KL) and factor analysis, is a linear (and unsupervised) model that under Gaussian assumptions derives global and orthogonal "features" that are now routinely referred to as eigenfaces. PCA lacks phase information and its use of global features refers to the fact that the support for each feature comes from the whole face image. Each eigenface represents one of the components or dimensions along which human faces are encoded (see Sect. 5.4). The eigenfaces were one of the first attempts made to define the face space and to compress the facial data into a small and compact biometric signature that can serve as a template for face recognition. A face is then approximated as a weighted combination of some ordered eigenfaces, with the weights found by projecting the face on the face space derived ahead of time using data independent of the faces whose identification or authentication one seeks. The set of weights constitutes the signature or template used later on for personal identification and authentication. Since relatively few eigenfaces are needed to create semblances of most people, this greatly reduces the amount of data that has to be stored in order to compare faces. Kirby and Sirovich were able to encode 115 Caucasian faces using only 40 eigenfaces. Turk and Pentland (1992) refined the techniques that had been pioneered by Kirby and Sirovich. Eigenspaces, were also defined locally as eigen features, to include eigen eyes, eigen mouth and eigen nose (Pentland et al., 1994). The eigenspaces and eigen features capture the global and local appearance of the face.

There are additional ways to define the face space, including Linear Discriminant Analysis (LDA) or (non-orthogonal and supervised) Fisher Linear Discriminant (FLD) (Etemad and Chellappa, 1997), Fisherfaces, which are LDA derived on eigen spaces (Belhumeuer et al., 1997), Independent Component Analysis (ICA) (Bartlett et al., 1998), and Evolutionary Pursuit (EP) (Liu and Wechsler, 2000), a projection pursuit method whose trajectory is traced by Genetic Algorithms (GA). Note that the eigenfaces are expressive rather than discriminative features and their usefulness should apply to face reconstruction rather than face identification and authentication, a role that is more suitable for LDA. Neural networks and Statistical Learning Theory (SLT) are another instantiation of the holistic approach. The WIZARD (Stonham, 1986), self-organizing feature maps (Kohonen, 1989), connectionism (Valentin et al., 1994), and support vector machines (SVM) are major examples for such an approach. There are direct connections between PCA and neural networks. Oja (1982) has shown that the hidden nodes for MLP span the same space as the one spanned by the leading eigen values for PCA. Such connections and their potential use for color compression are discussed later on (see Sect. 5.8). The (compact) hidden unit outputs were used as features by a second MLP (Golomb et al, 1991) for gender categorization.

The feature based approach gave way in the 1990s to a *structural approach*, similar in concept and scope with the earlier hybrid approach. The structural approach is now referred to as the *recognition-by-parts approach.* In addition to features there is a global structure linking the local features or parts. The Dynamic Link Architecture (DLA) (Lades et al., 1993) and its descendant, Elastic Bunch Graph matching (EBGM) (Wiskott, 1997), express the local features in terms of Gabor wavelets using the face as the underlying grid. The local features are linked within a graph with spring-like connections that define the face topography. The role for matching is to align between two (gallery and probe) graphs. EBGM bundles the features into bunches to allow for their variable appearance. Similarity between two faces corresponds to the cost paid for deformation or elastic alignment (Yuille, 1989). Another possible structure linking the 2D features for face recognition is the Hidden Markov Model (Samaria and Young, 1994). Similar to the structural approach in terms of plasticity is the *flexible appearance approach* pioneered by Lanitis et al. (1995). PCA is used to model the principal (inter- and intra- personal) modes of variation for both shape, i.e., face outline, and texture. Some methods can implement both the holistic and structural approach. As an example, Independent Component Analysis (ICA), depending on the architecture used, implements both the global (holistic) model and spatially localized features suitable for recognition-by-parts (Draper et al., 2003). Interestingly enough and what one would expect, the global features are best for face identification, with the local features best at recognizing facial expressions. The Local Features Analysis (LFA) (Penev and Atick, 1996) expands on standard PCA as it tries to fill in for some underlying structure. It does this by extracting sparsely distributed but topographically spaced local features from the global PCA modes. The grid used to index for the LFA kernels is reminiscent of the grid used by DLA and EBGM.

The most recent attempts to address face recognition are characteristic of *recognition-by-parts*. The overall encoding structure is referred to as a constellation (Heisele et al., 2003; Huang et al., 2003) (see Sect.9.7). The current parts or component-based recognition methods, known earlier on as aspect graphs or visual potentials (Koenderink and van Doorn, 1979), are suitable to handle partial occlusion and structural noise. The scope for pattern recognition, in general, and face recognition, in particular, has become much wider because training and/or testing can take place using incomplete or camouflaged/disguised patterns from single or multiple image sets. As we move from 2D stills to time-varying imagery and 3D, video tracking and recognition together with data fusion are the latest approaches for face recognition (see Chaps. 7, 8 and 9).

1.4 Road Map

This book, the first to comprehensively address the face recognition problem in its entirety, draws inspiration and gains new insights from complementary fields of endeavor, such as neurosciences, statistics, signal and image processing, computer vision, machine learning and pattern recognition, and statistical learning. The overall goal of the book is applied modern pattern recognition, with the understanding that the novel methods described are not restricted to faces but rather apply to any objects. The scope for pattern recognition considered is also much wider because both training and testing should take place using incomplete or camouflaged/disguised patterns drawn from single and/or multiple image sets. The various chapters treat topics related to how people represent, process and/or respond to the human face, modeling and prediction, representing the face space, identification and verification, face detection, tracking and recognition, 3D, data fusion, denial and deception under occlusion and disguise, performance evaluation and error analysis, and finally, competing security and privacy considerations.

The specific road map for the book is as follows. Ch. 2 brings forth the complementary dimensions of cognitive neurosciences, psychophysics, social sciences, and aesthetics and arts. They place the endeavor of face recognition in a multidisciplinary context, and provide the motivation and inspiration required to understand and advance the field of reliable face recognition. Perception, in general, and face recognition, in particular, requires training and reasoning or inference. This is discussed in Ch. 3 using the predictive learning framework, starting with the Bayesian approach, and continuing with connectionism or neural networks, statistical learning, and recent approaches such as transduction. The chapter ends with a comparative assessment of generative and discriminative approaches. Biometrics, in general, and face recognition, in particular, start with data capture. Towards that end, Ch. 4 considers sensing and enrollment, the standards required for proper biometric use and evaluation, and the compression means available to facilitate storage and processing. Human faces have to be represented before recognition can take place. Ch. 5 is involved with the means available to represent faces for fidelity purposes and enhanced discrimination. The notions and basics of the face space, scale space and invariance, are motivated and presented first. Specific subspace methods for face representation, e.g., eigenfaces and Fisherfaces, are then described and compared. Feature selection, caricatures, and kernel methods are among the methods proposed for more distinctive face representations that are expected to yield better performance.

Ch. 6 is involved with specific face recognition tasks such as verification and identification, watch list/surveillance, and selection and categorization. The chapter starts with the metrics available for measuring similarity between face representations, and their relative merits. Open set (face) recognition is then introduced and compared with closed set (face) recognition. Methods driven by transduction are described for implementing open set face recognition, and the recognition-by-parts strategy for face recognition is discussed in detail. Ch. 7 addresses the all encompassing problem of face in a crowd. It starts with eye and face detection, and continues with a thoroughly discussion on the related concepts of uncertainty, active learning, and evidence accumulation. Topics such as video break detection and key frame extraction, pose detection and manifolds, joint tracking and recognition, and subspace spatial-temporal analysis are described, and their specific benefits for face recognition using multiple image sets are explained. Ch. 8 considers and evaluates the use of 3D for face recognition. The topics discussed include sensing, the analysis by synthesis strategy of image interpretation, animation, and modeling and recognition in 3D using transformations and morphing to align enrolled and query data and measure their similarity. Ch. 9 is involved with data fusion. The motivation comes from the belief that more but independent sources of data are better at overcoming uncertainty and improving overall performance, or equivalently that the whole is more than the sum of its parts. Data fusion can involve multiple samples, multiple cues, multiple engines, several sensory channels, soft biometrics, or a combination thereof, using a voting scheme such as

AdaBoost. The chapter concludes with a description of how boosting and strangeness implement recognition-by-parts in a fashion characteristic of data fusion.

Biometrics can not assume that the personal signatures they have access to are complete and reliable. Towards that end, Ch. 10 considers means for deception and denial, e.g., occlusion and disguise, and human biases that face recognition algorithms are likely to exhibit. The chapter describes among others a number of counter measures to handle partial faces and camouflage, and determine if the biometric presented is alive. Ch. 10 concludes with a description of how adaptive and robust correlation filters can implement the recognition-by-parts strategy for handling occlusion and disguise. Ch. 11 considers augmented cognition to extend users' and face recognition engines' abilities in order to improve their performance and provide for graceful degradation. The chapter also discusses the important dimension of face expressions for face recognition and social communication. Chaps. 12 and 13 are involved with the important but closely related topics of performance evaluation and error analysis. The topics addressed in Ch. 12 include figures of merit, score normalization to account for different operating conditions, threshold settings and decision-making, choosing among competing face recognition engines, and the data bases available for training and testing face recognition algorithms. Ch. 13 discusses confidence intervals for performance indexes, fallacies that concern the effectiveness of mass screening and intrusion detection,, and anecdotal observations that clients are different with respect to the difficulty they present for being recognized and/or their ability to become imposters, and the means to handle such diversity of clients. The chapter concludes with a critical discussion of large-scale face recognition evaluations.

Ch. 14 expands on reliability to include security and privacy aspects. The chapter addresses the perceived threats and vulnerabilities, and the means to thwart them. The topics covered include aspects related to the diversity and uniqueness of biometrics, cryptographic means, steganography and digital watermarking for and using faces, anonymity and privacy and their preservation, and photofits to recall fleeting observances of biometric data. Ch. 15 is involved with expanding the scope for biometrics and adding to existing knowledge and practice. The rapid increase expected in biometric data volumes is not matched by a commensurate increase in the quantity or quality of data intensive scientific research tools. To meet such biometrics goals, the chapter introduces the idea of agent-based middleware, driven by machine learning, to automate the process of data search, query formulation, workflow configuration and service composition, and collaborative reuse, for enhanced biometric system design and performance. The Epilogue concludes the book with an overall assessment of challenges, and outlines promising R&D directions for their resolution.

2. The Human Face

The Face is the Soul of the Body
(Ludwig Wittgenstein)

Where does the human face come from and how did it evolve? More than 300 million years ago, the features of eyes, nostrils, and a hinged jaw have combined to create a face. Starting with prehistoric tetrapod creatures such as the (late Devonian) fish Panderichthys and the amphibian Acanthostega, the head/face complex has continued to be present and change ever since. "When the human fetus is five and a half weeks old and shaped like a bean, there appear from the underside three outgrowths. These bronchial arches develop, in fish, into gills. In mammals, these buds of tissue merge and mix to form our forehead, face, and throat, with the second arch moving upward to form the face" (Cole, 1998). As recounted by Cole, additional evolutionary changes that are responsible for the human face as we know it today include warm bloodedness that requires insulation and makes the skin softer, and a sense of vision that dominates over smell and touch and makes the eyes the center for the face. The facial hair went away and the jaws can now display face expressions. The new ways the food is ingested further shape and mold the facial bones, muscles, and the skull that harbors the human face.

Mammals are capable of recognizing each other's faces and of picking up important cues from the expressions imprinted on them. Sheep, in particular, excel and can remember other sheep faces after many years of separation (Kendrick et al., 2001). With the dominant sensory centers of hearing, smell, sight, and taste all located within the face framework, and the skull harboring the control center, the face commands much attention. When primates began to walk upright, the sense of sight becomes the most important of the senses. Body posture and vocalization gave way to face language and social intelligence. The face becomes the medium for sharing information and communicating within groups of increasing size. As human language and abstract reasoning have emerged only recently, a tantalizing question concerns the relation between face language and the early manifestations of human language. Face language, to some extent, is universal. Did the universal element of the face language transfer to universal elements for our primeval "mother" language?

The human face today is a collection of two bones (the skull and jaw) and 44 muscles, which are not attached to the bones. This enables a great liberty of movement that allows thousands of different facial expressions. Recognizing faces is absolutely central to our humanity; it is estimated that 50% of our brain function is devoted to vision, and that a lion's share of that goes for facial recognition. It is only natural that we endeavor to endow computers with the same ability. Another practical reason for computerized facial recognition was advanced by Gates (2004). She frames the problem of identification in historical perspective to bear not only on criminology but also on civil identification. In particular, she refers to Arendt (1973), who argued in *Origins of Totalitarianism* that "the claim to authority for determining who belongs and who does not is a central component of sovereignty." This has been "a particular preoccupation of modern states" and is even more so today. The human face also plays an important role in social interaction and communication that is crucial for realistic animation and video games. Last but not least, the face and its apparent beauty has been the focus for aesthetics and arts since the dawn of civilization. The beauty is either inner and hidden and reveals character, or outward and visible, and conveys physical appearance. The face is the messenger in both cases.

The book, mostly about the science and technology of automatic face recognition, brings forth in this chapter the complementary dimensions of cognitive neurosciences, psychophysics, social sciences, and aesthetics and arts. They provide both motivation and inspiration for how to advance the field of reliable face recognition.

2.1 Cognitive Neurosciences

From their early infancy, people have the ability to process faces. Babies have blurry vision at birth, but newborns can still discriminate, using external features, their mother's face from other female faces soon after birth. Infants can also differentiate both facial attractiveness and facial expressions. Is face recognition different from other categorization tasks regarding the processes the brain engages in? Not according to Gauthier and Logothetis (2000). There is neocortical and limbic cell selectivity in response to facial identity in the prefrontal cortex, to gaze direction in the superior temporal sulcus, to face expression in the amygdale, and to overall face configuration for most of the type of cells mentioned. Such selectivity, however, is not unique for face recognition. Gauthier and Logothetis found using neuro imaging on monkeys that preferential cell selectivity applies to "any arbitrary homogeneous class of artificial objects - which the animal has to individually learn, remember, and recognize again and again from a large number of distractors sharing a number of common features with the target. Faces are not "special" but rather the "default special" class in the primate recognition system."

Neurological patients and their syndromes provide a rich trove of information on how the brain is built and how it functions. Ramachandran (1998), like a sleuth always excited to "begin with a set of symptoms that seem bizarre and incomprehensible and then end up - at least in some cases - with an intellectual satisfying account in terms of the neural circuitry in the patient's brain," has worked for many years on the nature of phantom limbs, which are ghosts of arms and legs lost years before but still somehow remembered by the brain. The phantom limbs, according to Oliver Sacks who prefaced the book written by Ramachandran (with Sandra Blakeslee), serve as "experimental epistemology" and are explained by "reorganizations of body image in the sensory cortex." The sensory maps are thus not fixed but malleable. There is "nature" and there is "nurture." Maps can and do change as a result of injury. Neural connections are not fixed but rather plastic. There is a redundancy of connections and new paths can sprout, or even more intriguing paths can exists even for limbs missing since birth or never developed. A hand is lost and it becomes a phantom. Remapping takes place and the face, whose sensory area is right beside the hand, takes over the area previously allotted to the hand. Touching the face generates sensations in the phantom hand. The brain, modular with regard to functionality and localization, "doesn't hold all the answers." Genuine and spontaneous smiles are produced by basal ganglia but smiles on request come courtesy of the cortex.

The body surface is mapped on the surface of the brain behind the central sulcus. The Penfield's "sensory homunculus" is a visual rendering of how different parts of the body are mapped, and to what extent and where. The homunculus distorts the body and the size for different parts corresponding to their relative importance. The face and hand occupy "a disproportionately large share of the map." The explanation given, that "the area involved with lips and fingers takes up as much space as the area involved with the entire trunk of the body. This is presumably because your lips and fingers are highly sensitive to touch and are capable of very fine discrimination," is eminently plausible. The map is not continuous and "the face is not near the neck, where it should be, but is below the hand." The above findings suggest that varying rather than uniform resolution grids are used to represent the face. Self-Organization

Features Maps (SOFM) (see Sect. 5.6) could provide the mechanism used to allocate more grid space to the eyes, nose, and mouth at the expense of the cheeks. This involves competitive learning and Hebbian attractors that trade the real estate for the face representation between the facial areas and its landmarks according to their functional conspicuity and saliency.

While the processes by which the human brain recognizes faces are not fully understood, some clues are available from people affected by *Prosopagnosia*. The name of the disorder combines the Greek words for person and face (prosopon) and impairment (agnosia). More than failing to remember the names associated with the faces seen, individuals affected by prosopagnosia lose the ability to recognize faces [but still display autonomic covert recognition as measured by skin conductance responses], and lack any subjective sense of familiarity, even for their closest family members. Patients experience difficulty in tracking characters from TV shows and rely instead on non-facial information [similar to soft biometrics]. The disorder has to do with recall mechanisms and appears to be caused by an injury in the fusiform gyrus area of the brain, which involves the amygdala. Here, researchers have identified a *Fusiform Face Area* (FFA), an area specialized for face perception (Kanwisher et al., 1997). Yovel and Kanwisher (2004a) have shown, using fMRI studies of FFA, that face perception is domain rather than process specific. Subjects had to discriminate among pairs of upright or inverted faces or houses stimuli that differed in either the spatial distance among parts (configuration) or the shape of the parts. "The FFA showed a much higher response to faces than to houses, but no preference for the configuration task over the part task." Such findings are relevant to recognition-by-parts methods, which are compositional and structural in nature. The above findings appear to suggest that claims made on the existence of a generic dictionary of parts, e.g., geons (Biederman, 1987), are not warranted. The parts are rather different and according to the object they compose. They emerge as a result of competitive pressure encountered during discrimination tasks (see Sect. 9.7).

There are additional areas that appear to be involved in face processing. The result of evolution and functional differentiation, the areas are located in the posterior fusiform (PF), apparently a gateway to higher level processing including emotion, and in the middle temporal gyrus responsible for attention. Interesting also to note is that FFA lies in the "non retinotopic visual association cortex of the ventral visual processing stream" (Halgren et al., 1999). Canonical or configural configurations of face parts were found to trigger greater response vs. randomly rearranged parts within the face outline in the amygdala, superior temporal sulcus (STS), and FFA (Golarai et al., 2004). Deficits in configural processing could account for prosopagnosia (Duchaine et al., 2004). Face processing, however, is more than just configural. Face perception "engages a domain-specific system for processing both configural and part-based information about faces" (Yovel and Kanwisher, 2004b). This is needed to accommodate viewpoint or pose changes, occlusion and/or disguise, and temporal changes. Robust and steady part- or patch- based information can still identify a face despite missing and/or changed patches. Such information is behind the recent upsurge of "constellations" of parts for reliable face recognition (see Sects. 6.5 and 9.7).

It is the sub-ordinate rather than basic-level classification that appears to fire FFA. What about encoding for face recognition? "For stimuli such as faces, which are likely to be encountered by every member of the species, configural representations or [golden ratio] templates may be most effective because the basic stimulus configuration is invariant across the environments in which individuals may live. Thus the predictability of species-specific stimuli may allow for the creation through evolution of complex pattern recognition systems. These systems are tuned at birth but remain plastic through development" (Kanwisher and Moscovitch, 2000). The arguments listed above are relevant to basic-level face detection rather than sub-ordinate face identification. Liu, Harris et al. (2001) have MEG recordings to suggest that "face processing [indeed] proceeds through two stages: an initial stage of [basic] face categorization [after 100 ms], and a [70 ms] later stage at which the identity of the individual face is extracted."

Dissociations of face and object recognition in developmental prosopagnosia (Duchaine and Nakayama, 2005) support the hypothesis that "face and non-face recognition relies on separate mechanisms." Aware that "the acquisition of mature face perception skills is not complete until late adolescence" and that face recognition is a skill that has to be learned, the developmental aspect of the disorder refers to patients failing to develop the face recognition skills rather than acquiring the deficits as adults due to illness. What is the difference between object and face recognition? As recounted by Duchaine and Nakayama, "object recognition typically involves feature processing, but face recognition also involves holistic and configural processing. Holistic processing is characterized by the integration of facial information into a gestalt, whereas configural processing usually describe sensitivity to the precise spatial layout of the facial features." The development of specific mechanisms starts with external features for newborns, proceeds with internal features around eight weeks, and will continue with holistic and configural processing later on. Aspergers' syndrome is a mild form of autism characterized by an abnormally-sized amygdala. Patients are unable to recognize facial expressions, e.g., fear, and seem to analyze separate elements of a face more than the whole. Brain disorders, in general, and prosopagnosia and Aspergers syndromes, in particular, cannot be explained by any single cause. This makes a strong case for hybrid (local and global) approaches for face recognition that include (internal and external) features together with configural and holistic processing.

Another neurological disorder, reported by Dr. Ramachandran, is the Capgras' syndrome of misidentification, where the patient sees familiar and loved ones as impostors. Again we refer to Oliver Sacks and his explanation that there is "a clear neurological basis for the syndrome - the removal of the usual and crucial affective cues to recognition [leading to] affectless perceptions." Capgras' syndrome is explained by the damaged connections from the face processing areas in the temporal lobe to the limbic system. Patients report something like *she can't be my fiancée because I feel nothing* and record diminished galvanic skin response (GSR). Capgras' delusion suggests that there is more to face recognition than the face itself and that emotions play an important role. Antonio Damasio (1994) argues along similar lines for the role emotions play in rational thinking. Damasio's book *Descartes' Error* is a clever take on the famous *Je Pense donc Je suis* [I think therefore I exist]. Emotions are integral to existence. Some obvious implications for face recognition follow. Face expressions [of inner emotions] rather than being a handicap should help with face recognition as they are unique to each individual, e.g., the mysterious smile of Mona Lisa. Face expressions play also an imported role in interpreting human behaviors and augmented cognition (see Ch. 11). Video sequences rather than single still images for face recognition are thus required to capture the temporal dimension and the unique facial changes experienced. The neurological patient suffering from Capgars' syndrome remembers each face appearance episode but fails to link them into one category. The implications for face recognition are obvious. Multiple image frames are needed to search for the "glue" unique to each individual that makes the temporal sequence coherent.

Capgras' delusion is a mirror image of amnesia. A face is recognized but its authentication fails. Starting from the Capgras' delusion and within the framework of cognitive neuropsychiatry, Ellis and Lewis (2001) raise important epistemological questions on normal face recognition [circuitry] and the corresponding models of modal face recognition. The belief that specific cognitive functions are localized goes back to the anthropologist Paul Broca in the 19th century for whom an area of the frontal lobe is named. Today we witness a revolution in mapping brain functions to understand how the normal mind works. Neuroimaging is a "keystone for the growing field of cognitive neurosciences" (Culham, 2004). The growth of the field is indeed impressive. At the same time, Tulving, in an interview for *Cognitive Neuroscience* (Cooney et al., 2002), remarks that "what is badly needed now, with all these scanners whirring away, is an understanding of exactly what we are observing, and seeing, and measuring, and wondering about." Different but complementary neuroimaging technologies have become available, including electroencephalography (EEG), magnetic electroencephalography (MEG), magnetic resonance imaging (MRI), positron emission tomography (PET) and fMRI. The same technologies are also used to diagnose disorders and to monitor drug treatments. The underlying principle for neuroimaging is that matter, i.e., body or brain tissue, absorb energy at a resonant frequency and reemission or relaxation times that can be differentially measured (Lauterbur, 1973). Originally the devices were called NMR (nuclear magnetic resonance) but for public reassurance the name has changed later on to MRI (magnetic resonance imaging).

f(unctional)MRI, a standard MRI scanner, measures changes in blood-oxygenation-level dependent (BOLD) signals in response to variable magnetic fields. There is functional resolution, which relates neuronal activity and cognitive behavior, and there is also anatomical contrast, which differentiate between different properties of tissue. fMRI provides fine spatial and temporal resolution, 4 mm voxel size, and one second repeatability. This makes fMRI suitable for spatiotemporal event analysis using ICA (see Sect. 7.8). Differences in performance between attending vs. not attending to the stimulus were found higher for faces than for places in FFA, the apparent locus for face processing. This provides strong motivation for using attention mechanisms for face processing. As the temporal (ms) resolution for MEG is much higher than the one for fMRI, integral studies using fMRI and MEG are now being performed. "fMRI adaptation techniques hold excellent potential for evaluating the nature of the mental representations within an area. Both behavioral and fMRI adaptation work on the same principle: with continued stimulation, neurons show reduced responses" (Culham, 2004). The reduced response indicates the dimensions some brain area is sensitive to. As an example, the face selective FFA mentioned earlier yields lower response when subject to extended presentation of the same face compared to versions of different faces.

While suggesting that face recognition is modular [modal], Ellis and Lewis (2001) claim that the etiology behind Capgras' delusion "invokes a second stage at which autonomic [identification] attributions are made." Like Breen et al. (2000), Ellis and Lewis (2001) proposed (see Fig. 2.1) a modified dual-route model of face recognition and misidentification, where recognition takes place along the ventral route, and affective responses are provided by ventral limbic structures, especially the amygdala. In addition, they also proposed a second [integrative] facility that compares the conclusions of the two routes, which must be impaired for the delusion to take place. Abnormalities at locations A and B (see 2.1) are responsible for prosopagnosia and the Capgras' delusion, respectively. An abnormality at location C will not lead to delusions and would imply that damage at A or B could be circumvented. Furthermore, covert face recognition for patients affected by prosopagnosia is fractioned between autonomic and cognitive/behavioral recognition, e.g., face interference in terms of accuracy and latency, as measured by the skin conductance response (SCR) that was referred to earlier as GSR. The conceptual framework provided by the modal architecture has important implications for automatic face recognition. There are person identity nodes but there is also much scope for context and modularity and thus for multimodal and episodic recognition, and data (fusion and) integration.

Another fertile area of research is what information from faces is stored to make them recognizable. It has been noted that we see differences in our own ethnic group with greater ease than we do in other ethnic groups (the other race-effect or "they all look alike to me" syndrome.) Experiments show that we recognize a well done caricature faster than a photographic image, suggesting that the distinctive details of each face and their exaggeration or stereotyping lead to the correct match for identification (see Sect. 5.6).

2.2 Psychophysics

Psychophysics lies between perception and psychology. It is concerned with establishing qualitative and/or quantitative relations between physical stimulation and percep-

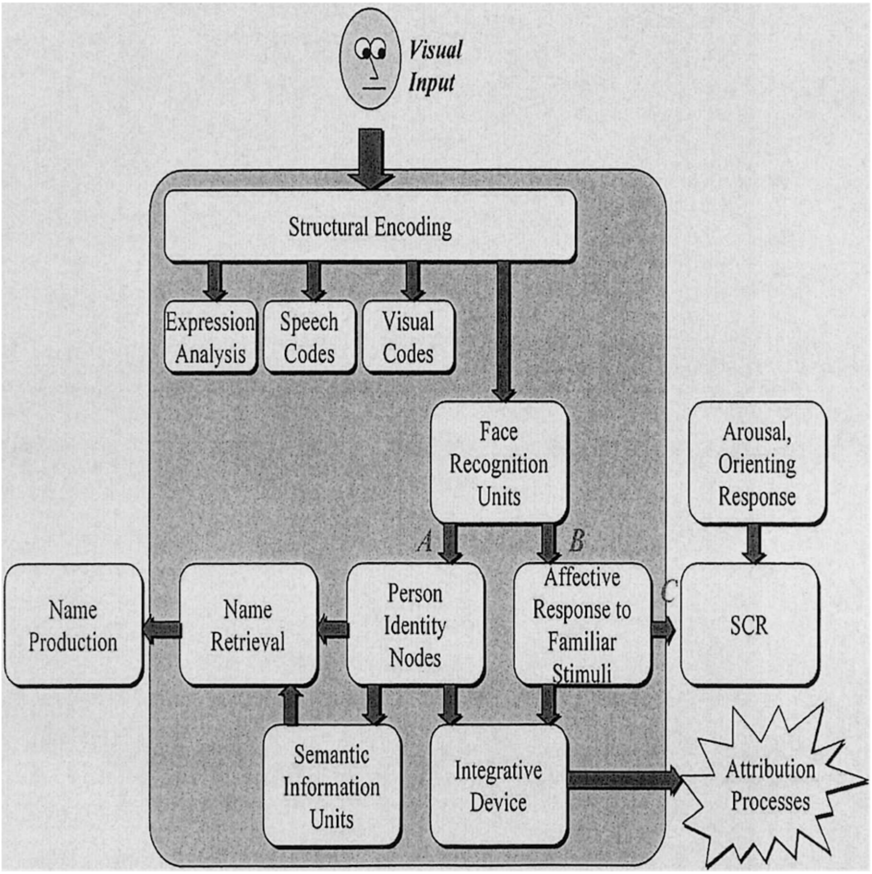


Fig. 2.1. Modal Brain Architecture for Face Recognition and Misidentification (Reprinted from *Trends in Cognitive Sciences*, Vol. 5, Ellis and Lewis, Capgras Delusion: A Window on Face Recognition, ©2001, with permission from Elsevier).