from:
N.J., Smelter, & P.B., Baltes (Eds.) (2001).
Encyclopedia of the Social and Behavioral Sciences.
London: Elsevier Science.

Article Title: ***Models of Face Recognition***

By: ***Alice J. O'Toole & Hervé Abdi***

Author Address: Alice O'Toole & Hervé Abdi, School of Human Development, MS: Gr.4.1, The University of Texas at Dallas, Richardson, TX 750833–0688, USA

Phone: 972 883 2065, ***fax:*** 972 883 2491 ***Date:*** 2001

E-mail: `otoole@utdallas.edu` `herve@utdallas.edu`

Abstract:
The human face provides us with valuable information about the identity of an individual, and about his/her sex, race, approximate age, and current mood. Models of face recognition encompass this broad range of tasks and include preprocessing steps aimed at aligning facial images/surfaces into a common coordinate system. Next, a process of feature extraction is implemented in many models by applying a principal component analysis (PCA) to a relatively raw perceptual encoding of faces. PCA produces a set of orthogonal feature axes that define a multidimensional face space into which individual faces can be projected. Recognition and identification models operate by assessing the distance of a target face to other known faces in the space. Categorization and facial expression analysis models are implemented generally with simple network algorithms that learn the mapping between faces represented by their coordinates in the space and a category membership (e.g., male). Current models are still limited in their ability to generalize recognition and categorization across moderate to large viewpoint changes.

Face Recognition Models

The *human face* is a highly meaningful stimulus that provides us with diverse information for adaptive social interaction with people. Our ability to recognize faces is remarkably accurate and long lasting. We are also able to categorize people along a number of visual dimensions including sex, race, and age and can readily interpret facial expression. The challenges associated with encoding and interpreting this information have become evident over the last two decades as psychologists, computer scientists, and cognitive scientists have endeavored to formulate *computational models* of these processes. The resultant models give insight into the complexity of the problems solved by the human brain in perceiving, representing, and remembering faces. In this article, computational approaches to modeling the *perception*, *categorization*, and *recognition* of human faces will be presented. The properties of the human face as a visual stimulus are described first, followed by definitions of the relevant tasks we perform with faces. The steps involved in modeling these tasks are reviewed next, and representative approaches for modeling individual tasks are discussed. Finally, the article closes with a few open questions in face recognition modeling.

1 The Human Face as a Visual Stimulus

The human face is a complex three-dimensional object defined by the structure of the skull and by the shape, texture, and pigmentation of the overlying skin and tissue. All faces share a basic set of features, (e.g., eyes, nose, and mouth, etc.) arranged in a well-defined configuration (eyes above the nose, etc). Individual faces comprise virtually limitless variations on this standard theme. To recognize an individual from a face, we must attend to the information that makes the face unique. To categorize a face we must extract and encode the information that a face shares with an entire category of faces (e.g., male faces), but which distinguishes the category from competing categories (e.g., female faces).

2 The Tasks

"Face recognition" models commonly encompass a range of tasks, including recognition, identification, verification, categorization, and the analysis of facial expression. *Face recognition* refers to the judgment of whether or not a particular face is "known". *Face identification* refers to the retrieval of information about the "owner" of the face, such as a name or context of encounter. *Face verification* refers to a decision about whether a particular face image belongs to a particular individual. Is this person John Doe? Face verification is a common goal of face algorithms developed for security systems.

3 Modeling: A Step by Step Approach

Face recognition models involve: a.) preprocessing algorithms to encode facial "features"; and b.) the application of this information to solve particular tasks.

3.1 Preprocessing Algorithms

3.1.1 Aligning Faces. All models that involve the analysis of a three-dimensional object from a two-dimensional image begin with the process of aligning the images into a common coordinate system. This facilitates feature extraction and comparison. Most current face recognition models operate effectively only with frontal images, tolerating only minimal changes in viewpoint. The alignment procedure employed in different models varies both in precision and in the degree of automaticity with which it is accomplished (i.e., by hand or by a computer algorithm). At the most basic level, alignment involves image translation, rotation, and scaling procedures implemented to assure that the eye levels are equivalent and that the centers of the foreheads correspond. More precise alignment is possible with *morphing* techniques that "warp" individual faces into the "average face" (Cameron and Craw 1991). To morph a face into another face (e.g., the average face), control points are located on the two faces (usually by hand). These consist of facial landmarks (e.g., corners of the eyes) and supplemental points (e.g., equally spaced points along the eyebrows). Using these points as a guide, each face is warped into the shape of the average face, yielding a correspondence of the control points across all faces. This alignment enables a separable encoding of the two-dimensional shape of the face and the image intensity information (see Section 3.1.2). Automated solutions to this *correspondence problem* have been implemented using all of the pixels or surface samples of the face rather than just a subset (Beymer and Poggio 1996; Blanz and Vetter 1999). These algorithms employ elaborated optic flow computations and work well on sets of faces for which correspondence is relatively easy to establish, (e.g., faces without hair that are pre-aligned with the translation method). Though difficult to achieve, when successful, complete alignment provides a powerful basis for synthesizing faces with arbitrary shapes and faces composed of intensity composites of other faces (Blanz and Vetter 1999).

A different approach to alignment is represented by the work of Lades et al. (1993) who developed a face recognition algorithm base on the dynamic link architecture. This algorithm combines alignment with identification. The model operates by placing a deformable grid over the target face, sampling the face at the grid vertices. The sampling is done with a series of oriented *Gabor wavelets*, designed to emulate the orientation specific neurons of visual cortex. The connectors between the vertices are allowed to deform elastically, enabling a re-sampling of the image until the best fit is obtained. The deformation parameters of this fit serve as the face representation, which is matched to the faces in the database to identify the best match.

3.1.2 Encoding and Representing Faces. The information in the aligned faces must be quantified in a way that enables recognition, identification, verification, categorization, and the analysis of expression in the model. What are the features of the face? We commonly think of the features of a face as its eyes, nose, and mouth. Descriptions of these features, such as those an eyewitness might provide, are inadequate for communicating enough information about an individual face to distinguish it from competing candidates. Geometrical measures, e.g., distance between eyes, have proved similarly inadequate (Laughery et al. 1981). More recent models have employed relatively raw perceptual codes, including roughly aligned images and three-dimensional surfaces, including pigmentation information. Another code, common since the advent of morphing technology, involves a two-component separable encoding of the two-dimensional face shape and the image intensities. The "shape" part of this code is defined as the deformation of the control points from the control points in the average face. The "shape-free" part of the code consists of a "shape standardized" two-dimensional array of image intensities created by warping an individual face into the shape of the average face.

In current computational and psychological models of face recognition, further analysis of these perceptual codes is carried out using a *principal component analysis* (PCA) (Sirovich and Kirby 1987; for a review see Valentin et al 1994). In the United States Government's tests of automatic face recognition algorithms between 1994-97, five of the seven algorithms tested used PCA. PCA is a statistical method for describing a set of correlated variables using a smaller number of uncorrelated or orthogonal variables. The uncorrelated variables are called eigenvectors or principal components (PCs), denoted $\mathbf{u}i$ and play the role of "features" for describing the faces. PCs can be considered features in the sense that any individual face, **f,** can be expressed as a linear combination of the PCs, $\acute{5}3\mathrm{i}wi\mathbf{u}I$, where the weights are the dot products, $wI = \mathbf{u}I\ \mathrm{T}\mathbf{f}$, between the faces and PCs. Because PCA is applied usually to images/surfaces, the PCs are also images/surfaces. Thus, individual faces can be synthesized as a linear combination of the PC images/surfaces. In geometrical terms, the PCA creates a multidimensional space in which the PCs define the axes of the space and individual faces are points in the space. The coordinates of a face in the space are the weights that specify the face's value on each PC feature. Note also, that three-layer *back propagation* networks can extract facial features similarly when they are trained to reconstruct faces through a bottleneck of hidden units. The hidden units of *these auto-encoders* have been shown to derive rotated versions of the PCs space (Cottrell et al. 1987).

PCA has appeal as a psychological model of face perception and memory for several reasons. First, it is consistent with psychological theories that posit a "face space" metaphor for human face memory (Valentine 1991). By this metaphor, faces can be thought of as points in a multidimensional space, with the distance between faces a measure of their similarity. At the center of the face space is the average or "prototype" face. The prototype, a central concept in psychological studies of face recognition, is invoked to explain the role of face typicality in predicting recognition performance. Typical faces, thought to be close to the prototype, are recognized less accurately than distinctive faces. This occurs presumably due to the greater density of faces close to the prototype, causing more confusion among typical faces than among distinctive faces. The prototype is also used as a reference face in creating automatic caricatures. *Caricatures* can be created by "moving a face" away from the average in the face space. This results in a more distinctive and recognizable version of the same face.

Second, the features that emerge from PCA are derived from the experience of the model. The role of experience in face recognition performance has been established perhaps most clearly in the phenomenon of the "other-race effect" — the finding that people recognize faces of their own-race more accurately than faces of other-races. This effect is predicted when the PCA is applied to a majority of faces of one race, and a smaller number of faces of other races. Because PCA derives its features from the statistical structure of the input faces, the resultant features are most appropriate for describing the majority race of faces. Consequently, less distinct encodings of minority race faces result because these faces are not well characterized by the features extracted primarily from the majority race of faces.

4 Tasks

4.1 Recognition

The quality of the stimulus representation determines the difficulty of the recognition or classification task. With a PCA-based representation, face recognition models can be implemented in a relatively simple way. A face is considered "known" when an image of the individual was part of the input used to create the PCA space. The most common "recognition" algorithm implements both recognition and identification. A target face is projected into the space and the distance to all other faces in the space is assessed. The nearest neighbor is chosen as the identity of the target face. Recognition can be implemented by setting a threshold distance, beyond which a target face is declared "unknown". An alternative and computationally more expedient algorithm for recognition assesses the representation error incurred by projecting the target face into the space. A threshold tolerance for error is used to determine whether a target face is known or novel.

4.2 Categorization

To categorize faces by sex, race, or age, individual exemplar faces must be assigned to different categories based on visually accessible facial features. Face *categorization* has been approached with supervised connectionist or *neural network classifiers*, such as the *perceptron* (see entry PERCEPTRONS). The models use examples to learn the mapping between face representations and categories. Numerous sex categorization models have been implemented and have been found to perform at or near human performance levels. A similarly structured race classifier has been implemented also, though the imbalance of experience most people have for the faces of different races must be implemented also to model human performance accurately. Finally, little work has been done on categorizing faces by age, though two complementary models of facial aging make use of morphing and caricaturing techniques, respectively. The former simulates aging by morphing individual faces toward the average of older faces (Burt and Perrett 1995). The latter simulates aging by caricaturing the three-dimensional head structure relative to a mean of young adult faces. Surprisingly, this results in an aged face (O'Toole et al. 1997).

4.3 Facial Expression Analysis

Models for categorizing faces by expression have been implemented in ways similar to sex and race classification models, but with somewhat less success. These models operate by mapping images of faces onto expression categories using supervised learning techniques. Representations have varied from aligned images, to PCAs of faces preprocessed by the Gabor wavelet filters described previously. Performance has been found to be well above chance, though still short of human performance on a similar task with similar stimuli. Facial expression analysis is currently a very active area of research and more published work on this problem is expected in the near future.

5.0 Open Questions

Despite the clear successes of face recognition models over the past two decades, the problem of recognizing faces from different viewpoints remains an unsolved challenge for models. Though part and parcel of the larger unsolved inverse optics problem of computer vision, the domain of faces may be more accessible due to the specific nature of face recognition as a within category problem. Some promising lines of research have begun and may soon yield new insights into this difficult problem (Edelman 1999).

Bibliography

Beymer D, Poggio T 1996 Image representations for visual learning. *Science* 272: 1905-1909.

Burt D M, Perrett D I 1995. Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society London B* 259*:* 137-143.

Blanz V, Vetter T 1999 A morphable model for the synthesis of 3D faces. *ACM SIGGRAPH Proceedings* 187-194.

Craw I, Cameron P 1991 Parameterizing images for recognition and reconstruction. In P. Mowforth (Ed.) *Proceedings of the British Machine Vision Conference* London: Springer-Verlag.

Cottrell G W, Munro P, Zipser D 1987 Learning internal representations of gray scale images: An example of extensional programming. *Proceedings of the 9th Annual Cognitive Science Society*, Lawrence Erlbaum

Edelman S 1999 *Representation and recognition in vision.* Cambridge: MIT Press.

Lades M, Vorbrueggen J C, Buhmann J, Lange J, von der Malsburg C, Wiskott R P, Konen, R W 1993 Distortion invariant face recognition in dynamic link architectures. *IEEE Transactions on Computers* 42: 300-311.

Laughery K, Rhodes B, Batten G 1981 Computer-guided recognition ad retrieval of facial images. In Davies G, Ellis H, Shepherd J (Eds*.) Perceiving and remembering faces.* London: UK Academic Press.

O'Toole A J, Vetter T, Volz H, Salter E M 1997 Three-dimensional caricatures of human heads : Distinctiveness and the perception of facial age. *Perception* 26: 719-732.

Sirovich L, Kirby M 1987 Low dimensional procedure for characterization of human faces. *Journal of he Optical Society of America A* 4: 518-519.

Valentin D, Abdi H, O'Toole A J, Cottrell G W 1994 Connectionist models of face processing: A survey. *Pattern Recognition* 27:1209-1230.

Valentine T 1991 A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology* 43A: 161-204.

Cross-referencing

2.2 Back propagation

2.2 Linear algebra neural networks

3.1.3 Object perception

Alice J. O'Toole and Herv8́e Abdi, The University of Texas at Dallas