A Review on Face Detection Methods

ARTICLE · FEBRUARY 2011

1 AUTHOR:

Dr Qaim Mehdi Rizvi
Qassim University

A Review on Face Detection Methods

Qaim Mehdi Rizvi	Prof. Bal Gopal Agarwal	Dr. Rizwan Beg
Asstt. Professor	DEAN (Academic)	HoD (CS)
CEST, Lucknow, India	IU, Lucknow, India	IU, Lucknow, India
[sirqaim@gmail.com]	[bgagarwal@yahoo.co.in]	[rizwanbeg@gmail.com]

Images containing faces are essential to intelligent vision-based human computer interaction, and research efforts in face processing include face recognition, face tracking, pose estimation, and expression recognition. As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention. This is evidenced by the emergence of face recognition conferences and systematic empirical evaluations of face recognition techniques. Numerous techniques have been developed to detect faces in a single image, and the purpose of this paper is to categorize and evaluate these algorithms.

1. Introduction:

With the ubiquity of new information technology and media, more effective and friendly methods for human computer interaction (HCI) are being developed which do not rely on traditional devices such as keyboards, mice, and displays. Furthermore, the ever decreasing price/ performance ratio of computing coupled with recent decreases in video image acquisition cost imply that computer vision systems can be deployed in desktop and embedded systems [1], [2], [3]. The rapidly expanding research in face processing is based on the premise that information about a user's identity, state, and intent can be extracted from images, and that computers can then react accordingly, e.g., by observing a person's facial expression. In the last five years, face and facial expression recognition have attracted much attention though they have been studied for more than 20 years by psychophysicists, neuroscientists, and engineers. Many research demonstrations and commercial applications have been developed from these efforts. A first step of any face processing system is detecting the locations in images where faces are present. However, face detection from a single image is a challenging task because of variability in scale, location, orientation (up-right, rotated), and pose (frontal, profile). Facial expression, occlusion, and lighting conditions also change the overall appearance of faces.

We now give a definition of face detection: Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face. The challenges associated with face detection can be attributed to the following factors:

- **Pose:** The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components:** Facial features such as beards, mustaches, and glasses may or may not be present and there is a great deal of variability among these components including shape, color, and size.
- **Facial expression:** The appearance of faces is directly affected by a person's facial expression.
- **Occlusion:** Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Image orientation:** Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions:** When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

There are many closely related problems of face detection. Face localization aims to determine the image position of a single face; this is a simplified detection problem with the assumption that an input image contains only one face [4], [5]. The goal of facial feature detection is to detect the presence and location of features, such as eyes, nose, nostrils, eyebrow, mouth, lips, ears, etc., with the assumption that there is only one face in an image [6], [7]. Face recognition or face identification compares an input image (probe) against a database (gallery) and reports a match, if any [8], [9], [10]. The purpose of face authentication is to verify the claim of the identity of an individual in an input image [11], [12], while face tracking methods continuously estimate the location and possibly the orientation of a face in an image sequence in real time [13], [14], [15]. Facial expression recognition concerns identifying the affective states (happy, sad, disgusted, etc.) of humans [16], [17]. Evidently, face detection is the first step in any automated system which solves the above problems. It is worth mentioning that many papers use the term "face detection," but the methods and the experimental results only show that a single face is localized in an input image. In this paper, we differentiate face detection from face localization since the latter is a simplified problem of the former. Meanwhile, we focus on face detection methods rather than tracking methods.

2. Face Detection using Single Image:

In this section, we review existing techniques to detect faces from a single intensity or color image. We classify single image detection methods into four categories; some methods clearly overlap category boundaries and are discussed at the end of this section.

- **Knowledge-based methods:** These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization.
- **Feature invariant approaches:** These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate faces. These methods are designed mainly for face localization.

- **Template matching methods:** Several standard patterns of a face are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are computed for detection. These methods have been used for both face localization and detection.
- **Appearance-based methods:** In contrast to template matching, the models (or templates) are learned from a set of training images which should capture the representative variability of facial appearance. These learned models are then used for detection. These methods are designed mainly for face detection.

2.1 Knowledge-Based Top-Down Methods:

In this approach, face detection methods are developed based on the rules derived from the researcher's knowledge of human faces. It is easy to come up with simple rules to describe the features of a face and their relationships. For example, a face often appears in an image with two eyes that are symmetric to each other, a nose, and a mouth. The relationships between features can be represented by their relative distances and positions. Facial features in an input image are extracted first, and face candidates are identified based on the coded rules. A verification process is usually applied to reduce false detections.

One problem with this approach is the difficulty in translating human knowledge into well-defined rules. If the rules are detailed (i.e., strict), they may fail to detect faces that do not pass all the rules. If the rules are too general, they may give many false positives. Moreover, it is difficult to extend this approach to detect faces in different poses since it is challenging to enumerate all possible cases. On the other hand, heuristics about faces work well in detecting frontal faces in uncluttered scenes.

Yang and Huang used a hierarchical knowledge-based method to detect faces [18]. Their system consists of three levels of rules. At the highest level, all possible face candidates are found by scanning a window over the input image and applying a set of rules at each location. The rules at a higher level are general descriptions of what a face looks like while the rules at lower levels rely on details of facial features. A multiresolution hierarchy of images is created by averaging and sub sampling, and an example is shown in Figure 2.1.

(a) (b) (c) (d)

Figure 2.1: (a) n=1, original image. (b) n=4, (C) n=8, (d) n=16. Original and corresponding low resolution images.

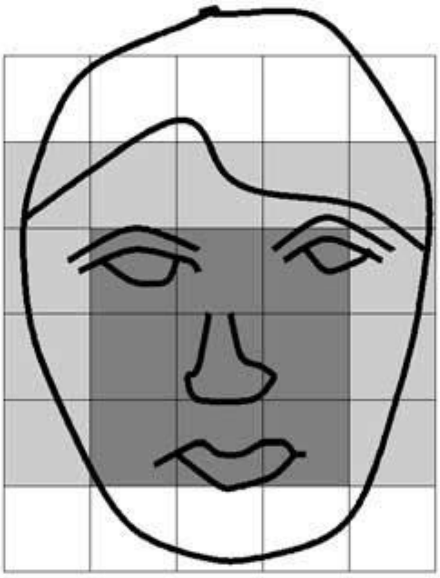


Figure 2.2: A typical face used in knowledge-based top-down methods: Rules are coded based on human knowledge about the characteristics (e.g., intensity distribution and difference) of the facial regions [18].

Kotropoulos and Pitas [19] presented a rule-based localization method which is similar to [20] and [18]. First, facial features are located with a projection method that Kanade successfully used to locate the boundary of a face [20]. Let I(x, y) be the intensity value of an m X n image at position (x,y), the horizontal and vertical projections of the image are defined as:

$$\mathbf{HI(x)} = \sum_{\mathbf{y=1}}^{\mathbf{n}} \mathbf{I(x,y)}$$
$$\mathbf{VI(y)} = \sum_{\mathbf{x=1}}^{\mathbf{n}} \mathbf{I(x,y)}$$

The horizontal profile of an input image is obtained first, and then the two local minima, determined by detecting abrupt changes in HI, are said to correspond to the left and right side of the head. Similarly, the vertical profile is obtained and the local minima are determined for the locations of mouth lips, nose tip, and eyes. Subsequently, eyebrow/eyes, nostrils/ nose, and the mouth detection rules are used to validate these candidates. The proposed method has been tested using a set of faces in frontal views extracted from the European ACTS M2VTS (Multi Modal Verification for Teleservices and Security applications) database [21] which contains video sequences of 37 different people. Each image sequence contains only one face in a uniform background. Their method provides correct face candidates in all tests.

2.2 Bottom-Up Feature-Based Methods:

In contrast to the knowledge-based top-down approach, researchers have been trying to find invariant features of faces for detection. The underlying assumption is based on the observation that humans can effortlessly detect faces and objects in different poses and lighting conditions and, so, there must exist properties or features which are invariant over these variabilities. Numerous methods have been proposed to first detect facial features and then to infer the presence of a face. Facial features such as eyebrows, eyes, nose, mouth, and hair-line are commonly extracted using edge detectors. Based on the extracted features, a statistical model is built to describe their relationships and to verify the existence of a face. One problem with these feature-based algorithms is that the image features can be severely corrupted due to illumination, noise, and occlusion. Feature boundaries can be weakened for faces, while shadows can cause numerous strong edges which together render perceptual grouping algorithms useless.

2.2.1 Facial Features:

Sirohey proposed a localization method to segment a face from a cluttered background for face identification [22]. It uses an edge map (Canny detector [23]) and heuristics to remove and group edges so that only the ones on the face contour are preserved. An ellipse is then fit to the boundary between the head region and the background. This algorithm achieves 80 percent accuracy on a database of 48 images with cluttered backgrounds.

Recently, Amit et al. presented a method for shape detection and applied it to detect frontal-view faces in still intensity images [24]. Detection follows two stages: focusing and intensive classification.

2.2.2 Skin Color:

Human skin color has been used and proven to be an effective feature in many applications from face detection to hand tracking. Although different people have different skin color, several studies have shown that the major difference lies largely between their intensity rather than their chrominance [7], [25], [26].

2.2.3 Texture:

Human faces have a distinct texture that can be used to separate them from different objects. Augusteijn and Skufca developed a method that infers the presence of a face through the identification of face-like textures [27]. The textures are computed using second-order statistical features (SGLD) [28] on subimages of 16 X 16 pixels. Three types of features are considered: skin, hair, and others. They used a cascade correlation neural network [29] for supervised classification of textures and a Kohonen self-organizing feature map [30] to form clusters for different texture classes.

2.3 Template Matching:

In template matching, a standard face pattern (usually frontal) is manually predefined or parameterized by a function. Given an input image, the correlation values with the standard patterns are computed for the face contour, eyes, nose, and mouth independently. The existence of a face is determined based on the correlation values. This approach has the advantage of being simple to implement. However, it has proven to be inadequate for face detection since it cannot effectively deal with variation in scale, pose, and shape. Multi-resolution, multi-scale, sub templates and deformable templates have subsequently been proposed to achieve scale and shape invariance.

2.3.1 Predefined Templates:

An early attempt to detect frontal faces in photographs is reported by Sakai et al. [31]. They used several subtemplates for the eyes, nose, mouth, and face contour to model a face. Each subtemplate is defined in terms of line segments. Lines in the input image are extracted based on greatest gradient change and then matched against the subtemplates. The correlations between subimages and contour templates are computed first to detect candidate locations of faces. Then, matching with the other subtemplates is performed at the candidate positions.

Sinha used a small set of spatial image invariants to describe the space of face patterns [32], [33]. His key insight for designing the invariant is that, while variations in illumination change the individual brightness of different parts of faces (such as eyes, cheeks, and forehead), the relative brightness of these parts remain largely unchanged. Determining pairwise ratios of the brightness of a few such regions and retaining just the "directions" of these ratios provides a robust invariant.

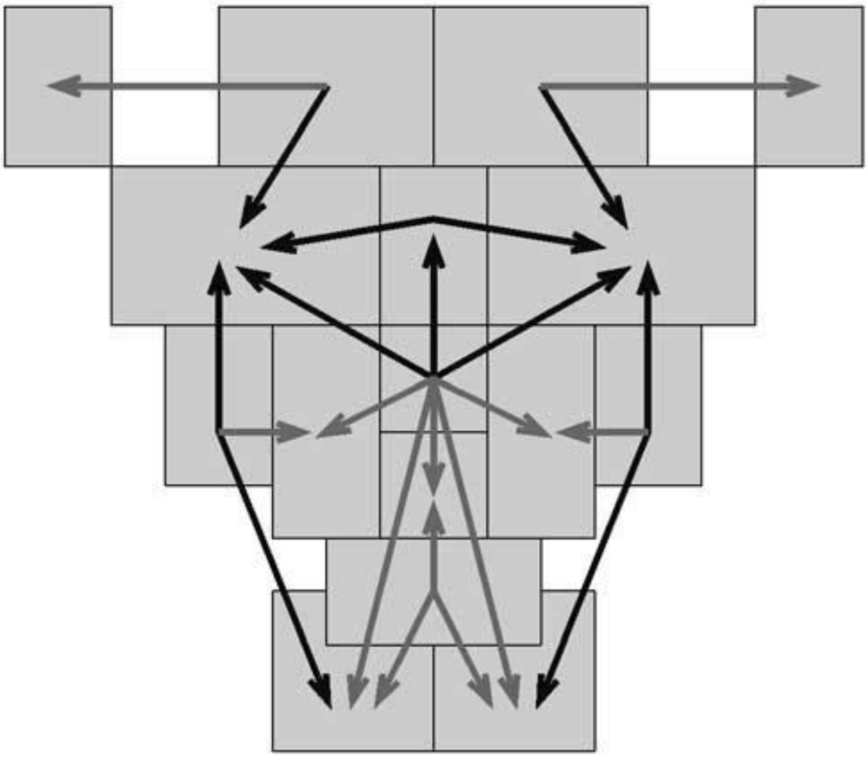


Figure 2.3: A 14x16 pixel ratio template for face localization based on Sinha method. The template is composed of 16 regions (the gray boxes) and 23 relations (shown by arrows) [34].

2.4 Appearance-Based Methods:

Contrasted to the template matching methods where templates are predefined by experts, the "templates" in appearance-based methods are learned from examples in images. In general, appearance-based methods rely on techniques from statistical analysis and machine learning to find the relevant characteristics of face and nonface images. The learned characteristics are in the form of distribution models or discriminant functions that are consequently used for face detection. Meanwhile, dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy.

Many appearance-based methods can be understood in a probabilistic framework. An image or feature vector derived from an image is viewed as a random variable x, and this random variable is characterized for faces and non faces by the class-conditional density functions p(x | face) and p(x | non face). Bayesian classification or maximum likelihood can be used to classify a candidate image location as face or nonface.

2.4.1 Distribution-Based Methods:

Sung and Poggio developed a distribution-based system for face detection [35], [36] which demonstrated how the distributions of image patterns from one object class can be learned from positive and negative examples (i.e., images) of that class. Their system consists of two components, distribution-based models for face/nonface patterns and a multilayer perceptron classifier. Each face and nonface example is first normalized and processed to a 19 X 19 pixel image and treated as a 361-dimensional vector or pattern. Next, the patterns are grouped into six faces and six nonfaces clusters using a modified k-means algorithm, as shown in Figure 2.4. Each cluster is represented as a multidimensional Gaussian function with a mean image and a covariance matrix. Figure 2.5 shows the distance measures in their method

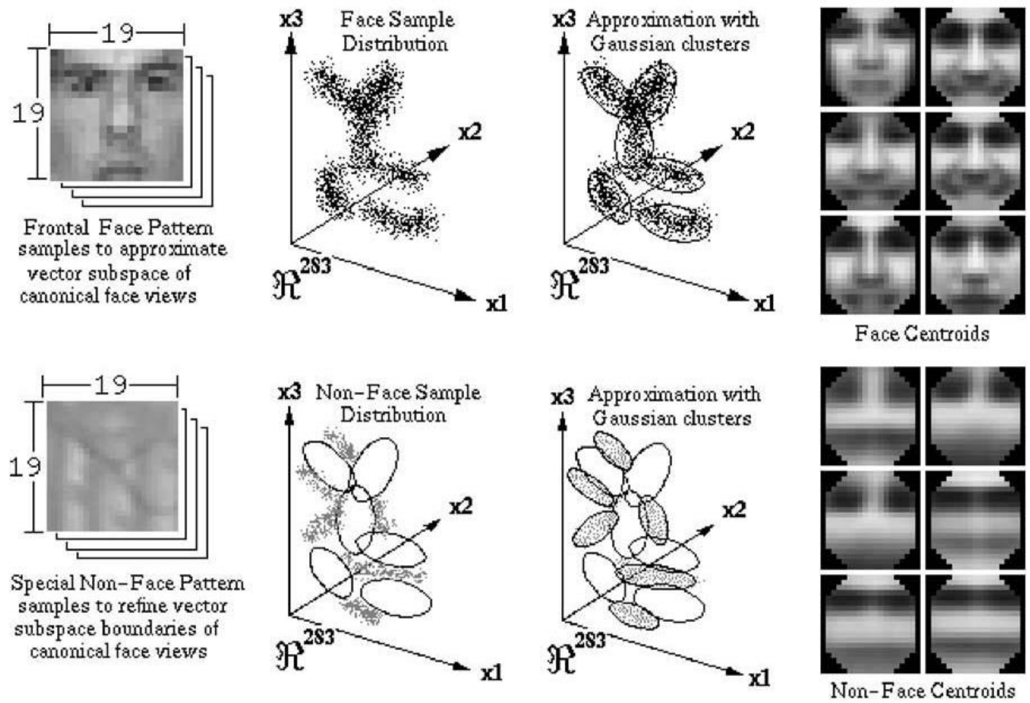
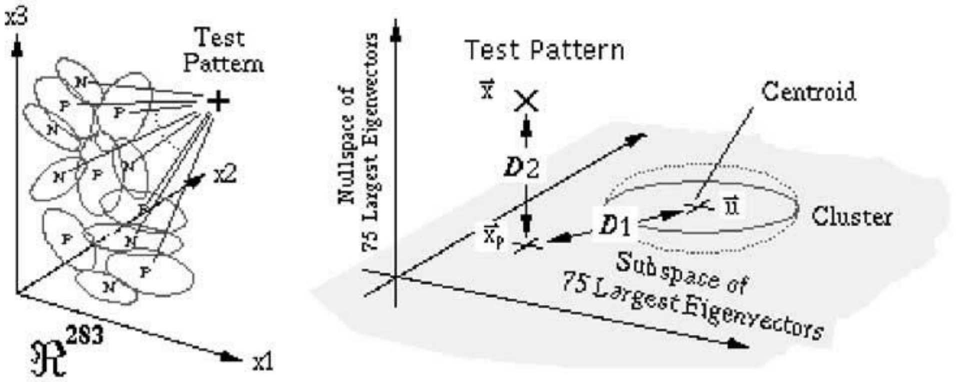


Figure 2.4: Face and nonface clusters used by Sung and Poggio [36]. Their method estimates density functions for face and nonface patterns using a set of Gaussians. The centers of these Gaussians are shown on the right (Courtesy of K.-K. Sung and T. Poggio).



(a) (b)

Figure 2.5: The distance measures used by Sung and Poggio [36]. Two distance metrics are computed between an input image pattern and the prototype clusters. (a) Given a test pattern, the distance between that image pattern and each cluster is computed. A set of 12 distances between the test pattern and the model's 12 cluster centroids. (b) Each distance measurement between the test pattern and a cluster centroid is a two-value distance metric. D1 is a Mahalanobis distance between the test pattern's projection and the cluster centroid in a subspace spanned by the cluster's 75 largest eigenvectors. D2 is the Euclidean distance between the test pattern and its projection in the subspace(Courtesy of K.-K. Sung and T. Poggio).

2.4.2 Support Vector Machines:

Support Vector Machines (SVMs) were first applied to face detection by Osuna et al. [37]. SVMs can be considered as a new paradigm to train polynomial function, neural networks, or radial basis function (RBF) classifiers. While most methods for training a classifier (e.g., Bayesian, neural networks, and RBF) are based on of minimizing the training error, i.e., empirical risk, SVMs operates on another induction principle, called structural risk minimization, which aims to minimize an upper bound on the expected generalization error. An SVM classifier is a linear classifier where the separating hyper-plane is chosen to minimize the expected classification error of the unseen test patterns. This optimal hyper-plane is defined by a weighted combination of a small subset of the training vectors, called support vectors.

2.4.3 Hidden Markov Model:

The underlying assumption of the Hidden Markov Model (HMM) is that patterns can be characterized as a parametric random process and that the parameters of this process can be estimated in a precise, well-defined manner. In developing an HMM for a pattern recognition problem, a number of hidden states need to be decided first to form a model. Then, one can train HMM to learn the transitional probability between states from the examples where each example is represented as a sequence of observations. The goal of training an HMM is to maximize the probability of observing the training data by adjusting the parameters in an HMM model with the standard Viterbi segmentation method and Baum-Welch algorithms [38]. After the HMM has been trained, the output probability of an observation determines the class to which it belongs.

HMM-based methods usually treat a face pattern as a sequence of observation vectors where each vector is a strip of pixels, as shown in Figure 2.6(a). During training and testing, an image is scanned in some order (usually from top to bottom) and an observation is taken as a block of pixels, as shown in Figure 2.6(a). For face patterns, the boundaries between strips of pixels are represented by probabilistic transitions between states, as shown in Figure 2.6(b), and the image data within a region is modeled by a multivariate Gaussian distribution. An observation sequence consists of all intensity values from each block. The output states correspond to the classes to which the observations belong. After the HMM has been trained, the output probability of an observation determines the class to which it belongs. HMMs have been applied to both face recognition and localization. Samaria [39] showed that the states of the HMM he trained corresponds to facial regions, as shown in Figure 2.6(b). In other words, one state is responsible for characterizing the observation vectors of human foreheads, and another state is responsible for characterizing the observation vectors of human eyes.

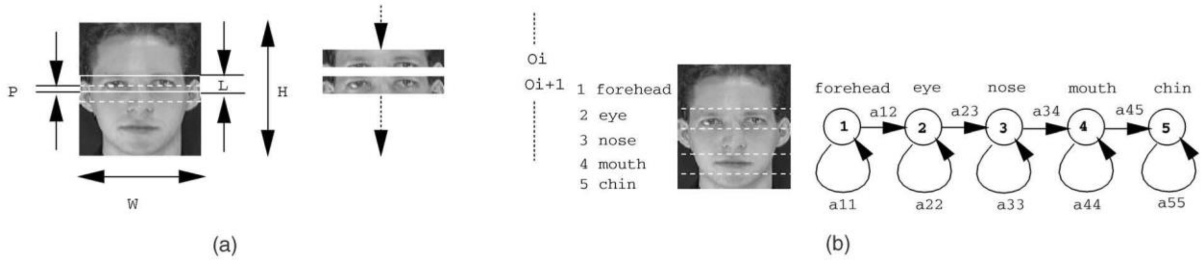


Figure 2.6: Hidden Markov model for face localization. (a) Observation vectors: To train an HMM, each face sample is converted to a sequence of observation vectors. Observation vectors are constructed from a window of W X L pixels. By scanning the window vertically with P pixels of overlap, an observation sequence is constructed. (b) Hidden states: When an HMM with five states is trained with sequences of observation vectors, the boundaries between states are shown in (b) [39].

3. Conclusion:

The scope of the considered techniques in evaluation is also important. In this survey, we discuss at least four different forms of the face detection problem:

- Localization in which there is a single face and the goal is to provide a suitable estimate of position; scale to be used as input for face recognition.
- In a cluttered monochrome scene, detect all faces.
- In color images, detect (localize) all faces.
- In a video sequence, detect and localize all faces.

An evaluation protocol should be carefully designed when assessing these different detection situations. It should be noted that there is a potential risk of using a universal though modest sized standard test set. As researchers develop new methods or “tweak” existing ones to get better performance on the test set, they engage in a subtle form of the unacceptable practice of “testing on the training set.” As a consequence, the latest methods may perform better against this hypothetical test set but not actually perform better in practice. This can be obviated by having a sufficiently large and representative universal test set. Alternatively, methods could be evaluated on a smaller test set if that test set is randomly chosen (generated) each time the method is evaluated.
Face detection is a challenging and interesting problem in and of itself. However, it can also be seen as a one of the few attempts at solving one of the grand challenges of computer vision, the recognition of object classes. The class of faces admits a great deal of shape,

color, and albedo variability due to differences in individuals, non-rigidity, facial hair, glasses, and makeup. Hence, face detection research confronts the full range of challenges found in general purpose, object class recognition. However, the class of faces also has very apparent regularities that are exploited by many heuristic or model-based methods or are readily "learned" in data-driven methods. One expects some regularity when defining classes in general, but they may not be so apparent. Finally, though faces have tremendous within class variability, face detection remains a two class recognition problem (face versus nonface).

4. References:

[1] A. Pentland, "Looking at People," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 107-119, Jan. 2000.

[2] A. Pentland, "Perceptual Intelligence," Comm. ACM, vol. 43, no. 3, pp. 35-44, 2000.

[3] A. Pentland and T. Choudhury, "Face Recognition for Smart Environments," IEEE Computer, pp. 50-55, 2000.

[4] K. Lam and H. Yan, "Fast Algorithm for Locating Head Boundaries," J. Electronic Imaging, vol. 3, no. 4, pp. 351-359, 1994.

[5] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, July 1997.

[6] I. Craw, D. Tock, and A. Bennett, "Finding Face Features," Proc. Second European Conf. Computer Vision, pp. 92-96, 1992.

[7] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating Faces and Facial Parts," Proc. First Int'l Workshop Automatic Face and Gesture Recognition, pp. 41-46, 1995.

[8] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," Proc. IEEE, vol. 83, no. 5, pp. 705-740, 1995.

[9] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," Pattern Recognition, vol. 25, no. 1, pp. 65-77, 1992.

[10] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[11] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal Face Authentication Uing Variants of Dynamic Link Matching Based on Mathematical Morphology," Proc. IEEE Int'l Conf. Image Processing, pp. 122-126, 1998.

[12] A. Tefas, C. Kotropoulos, and I. Pitas, "Variants of Dynamic Link Architecture Based on Mathematical Morphology for Frontal Face Authentication," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 814-819, 1998.

[13] J.L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 640-645, 1997.

[14] G.J. Edwards, C.J. Taylor, and T. Cootes, "Learning to Identify and Track Faces in Image Sequences." Proc. Sixth IEEE Int'l Conf. Computer Vision, pp. 317-322, 1998.

[15] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," Int'l J. Computer Vision, vol. 37, no. 2, pp. 175-185, 2000.

[16] I.A. Essa and A. Pentland, "Facial Expression Recognition Using a Dynamic Model and Motion Energy," Proc. Fifth IEEE Int'l Conf. Computer Vision, pp. 360-367, 1995.

[17] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, Oct. 2000.

[18] G. Yang and T. S. Huang, "Human Face Detection in Complex Background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994.

[19] C. Kotropoulos and I. Pitas, "Rule-Based Face Detection in Frontal Views," Proc. Int'l Conf. Acoustics, Speech and Signal Processing, vol. 4, pp. 2537-2540, 1997.

[20] T. Kanade, “Picture Processing by Computer Complex and Recognition of Human Faces,” PhD thesis, Kyoto Univ., 1973.
[21] S. Pigeon and L. Vandendrope, “The M2VTS Multimodal Face Database,” Proc. First Int’l Conf. Audio- and Video-Based Biometric Person Authentication, 1997.
[22] S.A. Sirohey, “Human Face Segmentation and Identification,” Technical Report CS-TR-3176, Univ. of Maryland, 1993.
[23] J. Canny, “A Computational Approach to Edge Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, June 1986.
[24] Y. Amit, D. Geman, and B. Jedynak, “Efficient Focusing and Face Detection,” Face Recognition: From Theory to Applications, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, and T.S. Huang, eds., vol. 163, pp. 124-156, 1998.
[25] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, “Multimodal System for Locating Heads and Faces,” Proc. Second Int’l Conf. Automatic Face and Gesture Recognition, pp. 88-93, 1996.
[26] J. Yang and A. Waibel, “A Real-Time Face Tracker,” Proc. Third Workshop Applications of Computer Vision, pp. 142-147, 1996.
[27] M.F. Augusteijn and T.L. Skujca, “Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology,” Proc. IEEE Conf. Neural Networks, pp. 392-398, 1993.
[28] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Texture Features for Image Classification,” IEEE Trans. Systems, Man, and Cybernetics, vol. 3, no. 6, pp. 610-621, 1973.
[29] S. Fahlman and C. Lebiere, “The Cascade-Correlation Learning Architecture,” Advances in Neural Information Processing Systems 2, D.S. Touretsky, ed., pp. 524-532, 1990.
[30] T. Kohonen, Self-Organization and Associative Memory. Springer 1989.
[31] T. Sakai, M. Nagao, and S. Fujibayashi, “Line Extraction and Pattern Detection in a Photograph,” Pattern Recognition, vol. 1, pp. 233-248, 1969.
[32] P. Sinha, “Object Recognition via Image Invariants: A Case Study,” Investigative Ophthalmology and Visual Science, vol. 35, no. 4, pp. 1735-1740, 1994.
[33] P. Sinha, “Processing and Recognizing 3D Forms,” PhD thesis, Massachusetts Inst. Of Technology, 1995.
[34] B. Scassellati, “Eye Finding via Face Detection for a Foevated, Active Vision System,” Proc. 15th Nat’l Conf. Artificial Intelligence, 1998.
[35] K.-K. Sung, “Learning and Example Selection for Object and Pattern Detection,” PhD thesis, Massachusetts Inst. of Technology, 1996.
[36] K.-K. Sung and T. Poggio, “Example-Based Learning for View- Based Human Face Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51, Jan. 1998.
[37] E. Osuna, R. Freund, and F. Girosi, “Training Support Vector Machines: An Application to Face Detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997.
[38] L.R. Rabiner and B.-H. Jung, Fundamentals of Speech Recognition. Prentice Hall, 1993.
[39] F.S. Samaria, “Face Recognition Using Hidden Markov Models,” PhD thesis, Univ. of Cambridge, 1994.