Artificial Intelligence Review (2019) 52:927–948
https://doi.org/10.1007/s10462-018-9650-2

Face detection techniques: a review

Ashu Kumar[1] · Amandeep Kaur[2] · Munish Kumar[3]

Published online: 4 August 2018

Abstract
With the marvelous increase in video and image database there is an incredible need of automatic understanding and examination of information by the intelligent systems as manually it is getting to be plainly distant. Face plays a major role in social intercourse for conveying identity and feelings of a person. Human beings have not tremendous ability to identify different faces than machines. So, automatic face detection system plays an important role in face recognition, facial expression recognition, head-pose estimation, human–computer interaction etc. Face detection is a computer technology that determines the location and size of a human face in a digital image. Face detection has been a standout amongst topics in the computer vision literature. This paper presents a comprehensive survey of various techniques explored for face detection in digital images. Different challenges and applications of face detection are also presented in this paper. At the end, different standard databases for face detection are also given with their features. Furthermore, we organize special discussions on the practical aspects towards the development of a robust face detection system and conclude this paper with several promising directions for future research.

Keywords Face detection · Eigen faces · PCA · Feature analysis

1 Introduction

With the rapid increase of computational powers and accessibility of innovative sensing, analysis and rendering equipment and technologies, computers are becoming more and more intelligent. Many research projects and commercial products have demonstrated the capa-

✉ Munish Kumar
munishcse@gmail.com

Ashu Kumar
ashu_sa@pbi.ac.in

Amandeep Kaur
aman_k2007@hotmail.com

[1] Department of Computer Science, Punjabi University, Patiala, Punjab, India

[2] Centre of Computer Science and Technology, Central University of Punjab, Bathinda, Punjab, India

[3] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

bility of a computer to interact with humans in a natural way by looking at people through cameras, listening to people through microphones, understanding these inputs, and reacting to people in a friendly manner. One of the fundamental techniques that enable such natural Human–Computer Interaction (HCI) is face detection. Face detection is the step stone to all facial analysis algorithms, including the face alignment, face modelling, face relighting, face recognition, face verification/authentication, head pose tracking, facial expression tracking/recognition, gender/age recognition, and many more. So, computers can understand face clearly, after that they begin to truly understand people's thoughts and intentions. Given a digital image, the primary goal of face detection is to determine whether or not there are any faces in the image. This appears as a trivial task for human beings, but it is a very challenging task for computers, and has been one of the top studied research topics in the past few decades. The difficulty associated with face detection can be attributed to many variations in scale, location, orientation (in-plane rotation), pose (out-of-plane rotation), facial expression, lighting conditions, occlusions, etc. A lot of reports are available for face detection in the literature. The field of face detection has made considerable progress in the past decade.

Mukherjee et al. (2017) have discussed the formulation for both the methods, i.e., using hand-crafted features followed by training a simple classifier and an entirely modern approach of learning features from data using neural networks. Ren et al. (2017) have presented a method for real time detection and tracking of the human face. The proposed method combines the Convolution Neural Network detection and the Kalman filter tracking. Convolution Neural Network is used to detect the face in the video, which is more accurate than traditional detection method. When the face is largely deflected or severely occluded, Kalman filter tracking is utilized to predict the face position. They try to increase the face detection rate, while meeting the real time requirements. Luo et al. (2018) have suggested deep cascaded detection method that iteratively exploits bounding-box regression, a localization technique, to approach the detection of potential faces in images. They also consider the inherent correlation of classification and bounding-box regression and exploit it to further increase overall performance. Their method leverages cascaded architecture with three stages of carefully designed deep convolutional networks to predict the existence of faces. TensorFlow is a machine learning system that operates on large scale and in heterogeneous environments. Tensor-Flow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. It maps the nodes of a dataflow graph across many machines in a cluster, and within a machine across multiple computing devices, including multicore CPUs, general purpose GPUs, and custom-designed ASICs known as Tensor Processing Units (TPUs). This architecture gives flexibility to the application developer: whereas in previous "parameter server" designs the management of shared state is built into the system. TensorFlow enables developers to experiment with novel optimizations and training algorithms. TensorFlow supports a variety of applications, with a focus on training and inference on deep neural networks. Several Google services use TensorFlow in production. They describe the TensorFlow dataflow model and demonstrate the compelling performance that Tensor-Flow achieves for several real-world applications (Abadi et al. 2016) (Figs. 1, 2).

In this paper, authors have presented a comprehensive survey on the recently used techniques for face detection from digital images. The rest of the paper is organized as follows. Section 2 gives an overview of challenges to face detection. The applications of face detection are presented in Sect. 3. Section 4, present various techniques for face detection. Existing standard databases of face detection are discussed in Sect. 5. Conclusion and future directions are given in Sect. 6.

Fig. 1 A sample of faces

Fig. 2 Detected faces

2 Challenges in face detection

Challenges in face detection, are the reasons which reduce the accuracy and detection rate of face detection. These challenges are complex background, too many faces in images, odd expressions, illuminations, less resolution, face occlusion, skin color, distance and orientation etc. (Figure 3).

- *Odd expressions* Human face in an image may have odd expressions unlike normal, which is challenge for face detection.
- *Face occlusion* Face occlusion is hiding face by any object. It may be glasses, scarf, hand, hairs, hats and any other object etc. It also reduces the face detection rate.
- *Illuminations* Lighting effects may not be uniform in the image. Some part of the image may have very high illumination and other may have very low illumination.
- *Complex background* Complex background means a lot of objects presents in the image, which reduces the accuracy and rate of face detection.
- *Too many faces in the image* It means image contains too many human faces, which is challenge for face detection.
- *Less resolution* Resolution of image may be very poor, which is also challenging for face detection.
- *Skin color* Skin-color changes with geographical locations. Skin color of Chinese is different from African and skin-color of African is different from American and so on. Changing skin-color is also challenging for face detection.

Fig. 3 Various categories of challenges for face detection

- *Distance* Too much distance between camera and human face may reduce the detection rate of human faces in image.
- *Orientation* Face orientation is the pose of face with an angle. It also reduces the accuracy and detection rate of face detection.

3 Applications of face detection system

- *Gender classification* Gender information can be found from human being image.
- *Document control and access control* Control can be imposed to document access with face identification system.
- *Human computer interaction system* It is design and use of computer technology, focusing particularly on the interfaces between users and computers.
- *Biometric attendance* It is system of taking attendance of people by their finger prints or face etc.
- *Photography* Some recent digital cameras use face detection for autofocus. Face detection is also useful for selecting regions of interest in photo slideshows.
- *Facial feature extraction* Facial features like nose, eyes, mouth, skin-color etc. can be extracted from image.
- *Face recognition* A facial recognition system is a process of identifying or verifying a person from a digital image or a video frame. One of the ways to do this is by comparing selected facial features from the image and a facial database. It is typically used in security systems.
- *Marketing* Face detection is gaining the interest of marketers. A webcam can be integrated into a television and detect any face that walks by. The system then calculates the race, gender, and age range of the face. Once the information is collected, a series of advertisements can be played that is specific towards the detected race/gender/age.

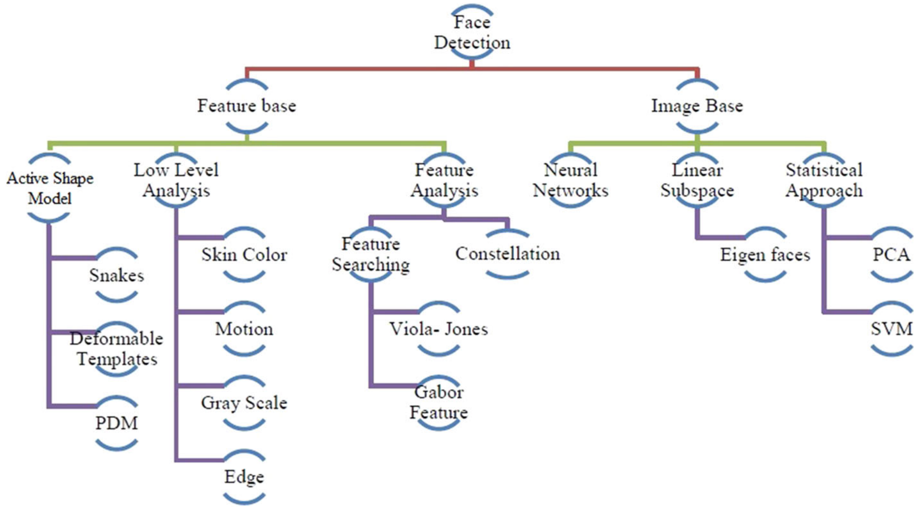


Fig. 4 Different techniques for face detection

4 Face detection techniques

Face detection is a computer technology that determines the location and size of a human face in the digital image. The facial features are detected and any other objects like trees, buildings and bodies are ignored from the digital image. It can be regarded as a specific case of object-class detection, where the task is finding the location and sizes of all objects in an image that belongs to a given class. Face detection, can be seen as a more general case of face localization. In face localization, the task is to identify the locations and sizes of a known number of faces (usually one). Basically, there are two types of approaches to detect facial part in the given digital image i.e. feature based and image based approach. Feature based approach tries to extract features of the image and match it against the knowledge of the facial features. While image based approach tries to get the best match between training and testing images. The following methods are commonly used to detect the faces from a still image or a video sequence (Fig. 4).

4.1 Features based approaches

4.1.1 Active shape model

Active Shape Model (ASM) focus on complex non-rigid features like actual physical and higher level appearance of features. Main aim of ASM is automatically locating landmark points that define the shape of any statistically modelled object in an image. For examples, in an image of human being face, extracted features such as the eyes, lips, nose, mouth and eyebrows. The training stage of an ASM involves the building of a statistical facial model containing images with manually annotated landmarks. ASMs is classified into three groups i.e. Snakes, Point Distribution Model (PDM) and deformable templates.

Snakes The first type uses a generic active contour called snakes (Kass et al. 1988). Snakes are used to identify head boundaries. In order to achieve the task, a snake is first initialized at the proximity around a head boundary. It then looks onto nearby edges and subsequently assumes the shape of the head. The evolution of a snake is achieved by minimizing an energy function, Esnake (analogy with physical systems), denoted as:

$$\text{Esnake} = \text{Einternal} + \text{Eexternal}$$

where Einternal and Eexternal are internal and external energy functions.

Internal energy is the part that depends on the intrinsic properties of the snake and defines its natural evolution. The typical natural evolution in snakes is shrinking or expanding. The external energy counteracts the internal energy and enables the contours to deviate from the natural evolution and eventually assume the shape of nearby features-the head boundary at a state of equilibria. Two main considerations for forming snakes i.e. selection of energy terms and energy minimization. Elastic energy (Erik and Low 2001) is used commonly as internal energy. Internal energy is varying the distance between control points on the snake, through which we get contour, an elastic-band characteristic that causes it to shrink or expand. On other side external energy relay on image features. Energy minimization process is done by optimization techniques such as the steepest gradient descent, which needs the highest computations. Fast iteration methods by greedy algorithms are also used. Snakes have some demerits like contour often become trapped onto false image features and another one is that snakes are not suitable in extracting non convex features.

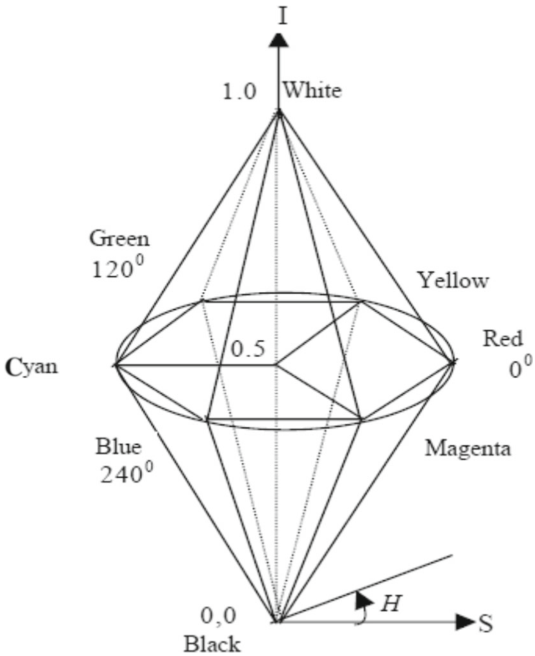
Point distribution model Point Distribution Model (PDM) was developed independent of computerized image analysis, and developed statistical models of shape (Erik and Low 2001). The idea is that once one can represent shapes as vectors, after that they can apply standard statistical methods to them just like any other multivariate object. These models learn allowable constellations of shape points from training examples and use principal components to build model, known as Point Distribution Model (PDM). These have been used in diverse ways, for example for categorizing Iron Age broaches. The first parametric statistical shape model for image analysis based on principal components of inter-landmark distances were presented (Cootes et al. 1992). Based on this approach, they released a series of papers that cumulated in what we call the classical Active Shape Model.

Deformable templates Deformable templates take into account a priori of facial features to improve the performance of snakes (Yuille et al. 1992). Locating a facial feature boundary does not constitute an easy task because the local evidence of facial edges is difficult to organize into a sensible global entity using generic contours. The low brightness contrast around some of these features also makes the edge detection process problematic. Concept of snakes is a step further by incorporating global information of the eye takes to improve the reliability of the extraction process. Deformable templates approaches are designed to solve this problem. Deformation is based on narrow valley, edge, peak, and brightness. Other than face boundary, salient feature (eyes, nose, mouth and eyebrows) extraction is a great challenge of face recognition.

$$\text{E} = \text{Ev} + \text{Ee} + \text{Ep} + \text{Ei} + \text{Einternal}$$

where Ev, Ee, Ep, Ei, Einternal are external energy due to valley, edges, peak and image brightness and internal energy.

Fig. 5 Double cone model



4.1.2 Low level analysis

Skin color base Color is an important feature of human faces. Using skin color as a feature for tracking a face has several advantages. Color processing is much faster than other facial features. Under certain lighting conditions, color is orientation invariant. This property makes motion estimation much easier because only a translation model is needed for motion estimation. Tracking human faces using color as a feature has several problems like the color representation of a face obtained by a camera is influenced by many factors like, ambient light, object movement, etc. suggested simplest skin-color algorithms for detecting skin pixels have been suggested by Crowley and Coutaz in (1997). The perceived human color varies as a function of the relative direction to the illumination. Pixels for skin region can be detected using a normalized color histogram, and can be normalized for changes in intensity on dividing by luminance (Fig. 5).

Conversion of an [R, G, B] vector into an [r, g] vector of normalized color which provides a fast processing of skin detection. This algorithm fails when there are some more skin regions like legs, arms, etc. Skin color classification algorithm with YCbCr color space is also introduced. Few researchers have noticed that pixels belonging to skin regions have similar Cb and Cr values. So, the thresholds be chosen as [Cr1, Cr2] and [Cb1, Cb2], a pixel is classified to have skin tone if the values [Cr, Cb] fall within the thresholds. The skin color distribution gives the face portion in the color image. This algorithm is also having the constraint that the image should be having only face as the skin region. A color predicate in HSV color space to separate skin regions from background has been defined (Kjeldsen and Kender 1996). Skin color classification in HSI color space is the same as YCbCr color space but here the responsible values are hue (H) and saturation (S). Similar to above the threshold be chosen as [H1, S1] and [H2, S2], and a pixel is classified to have skin tone if the values [H,

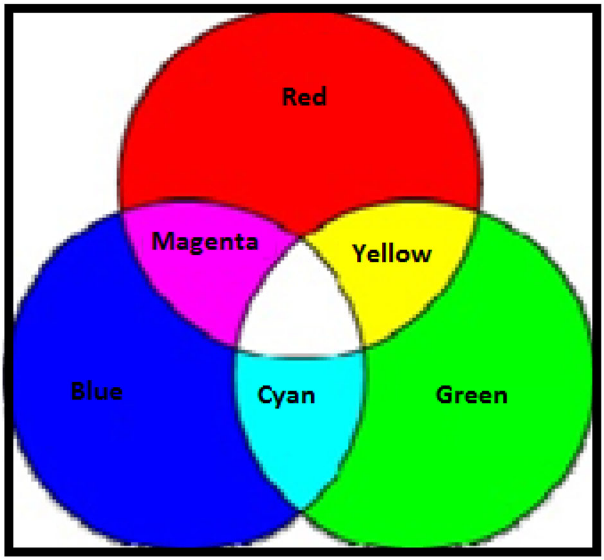


Fig. 6 RGB color model

S] fall within the threshold and this distribution gives the localized face image. Generally, three different face detection algorithms are available based on RGB, YCbCr, and HIS color space models. For implementation of these algorithms there are basically three main steps are required as the following:

- Classify the skin region in the color space,
- Apply threshold to mask the skin region and
- Draw bounding box to extract the face image.

RGB color model RGB colors are specified in terms of three primary colors i.e. Red (R), Green (G), and Blue (B) (Fig. 6). In RGB color space, a normalized color histogram is used to detect the pixels of skin color of an image and can be further normalized for changes in intensity on dividing by luminance. This localizes and detects the face (Subban and Mishra 2012). It is the basic color model and all other color models are derived from it. The RGB color model is light sensitive. In comparison with other color models such as YCbCr or HSI it has a major drawback that it cannot separate clearly the mere color (Chroma) and the intensity of a pixel, so it is sometimes difficult to distinguish skin colored regions. These factors contribute to the less favourable of RGB. It is widely used color space for processing and storing digital images because of the fact that chrominance and luminance components are mixed in RGB; it is not widely used in skin detection algorithm. Dhivakar et al. (2015) have proposed a method which consists of two main parts as detection of faces and then recognizing the detected faces. In detection step, skin color segmentation with thresholding skin color model combined with AdaBoost algorithm is used, which is fast and also more accurate in detecting the faces. Also, a series of morphological operators is used to improve the face detection performance. Recognition part consists of three steps: Gabor feature extraction, dimension reduction and feature selection using PCA, and KNN based classification. The system is robust enough to detect faces in different lighting conditions, scales, poses, and skin colors (Fig. 7).

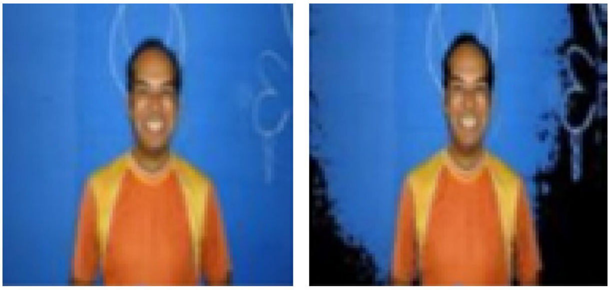


Fig. 7 **a** Original image and **b** processed image with RGB model

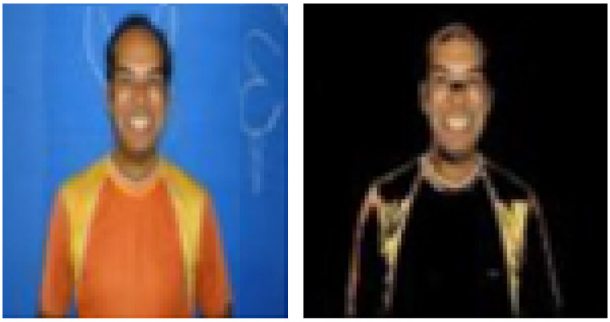


Fig. 8 **a** Original image and **b** processed image with HSV model

HSV color model HSV colors are specified in terms of Hue (H), Saturation (S) and Intensity value (V) which are the three attributes. Hue refers to the color of red, green, blue and yellow having the range of 0–360. Saturation means purity of the color and takes the value of 0–100% whereas Value refers to the brightness of color and provides the achromatic idea of the color (Hashem 2009). From this color space, H and S will provide the essential information about the skin color (Fig. 8). The skin color pixel should satisfy the following condition:

$$0 \leq \mathrm{H} \leq 0.25$$
$$0.15 \leq \mathrm{S} \leq 0.9$$

The transformation between HSV and RGB is non-linear (Crowley and Coutaz 1997). In the HSV color model, Hue (H) is not reliable for the discrimination task when the saturation is low. But, where color description is considering HSV color model rather than RGB model.

YCbCr color model YCbCr Color model is specified in terms of luminance (Y channel) and chrominance (Cb and Cr channels). It segments the image into a luminous component and chrominance components. In YCbCr color model, the distribution of the skin areas is consistent across different races in the Cb and Cr color spaces (Zhu et al. 2012a, b). As RGB color model is light sensitive so to improve the performance of skin color clustering, YCbCr color model is used. Its chrominance components are almost independent of luminance and there is non-linear relationship between chrominance (Cb, Cr) and luminance(Y) of the skin color in the high and low luminance region (Agui et al. 1992). Range of Y lies between 16

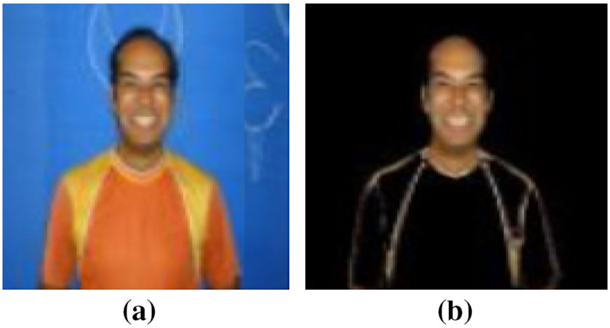
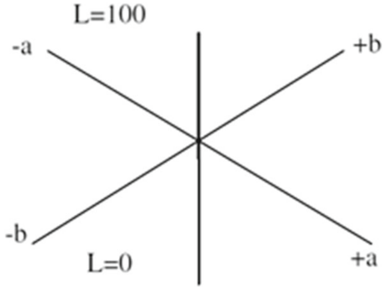


Fig. 9 **a** Original image and **b** processed image with YCbCr model

Fig. 10 CIELAB color model



and 235 where 16 for black and 235 for white whereas Cb and Cr are scaled in the range of 16–240. The main advantage of YCbCr color model is that the influence of luminosity can be removed during processing of an image. Different plots for Y, Cb and Cr values for face and non-face pixels were plotted using the reference images and studied to find the range of Y, Cb and Cr values for the face pixels (Fig. 9).

CIELAB color model In 1976, the CIE (International Commission on Illumination) recommended the CIEL*a*b* or CIELAB, color scale for use. It provides a standard, approximately uniform color scale which could be used by everyone so that the color values can be easily compared. This color model is designed to approximate perceptually uniform Color spaces (UCSs). It is related to the RGB color space through a highly nonlinear transformation. Examples of similar color spaces are CIE-Luv and Farnsworth UCS (Zou and Kamata 2010). It has three axes in it two are color axes and the third is lightness (Fig. 10).

Where L indicates lightness, $+a$ and $-a$ indicates amount of green and red color, respectively, $+b$ and $-b$ indicates amount of yellow and blue color, respectively. Here, maximum value of L is 100 which represent a perfect reflecting diffuser (white color) and the minimum value for L is 0 which represent *s* black color. Axes a and b do not have any specific numerical value (Fig. 11).

Comparative study of RGB versus HSV and RGB versus YCbCr color model

- RGB versus HSV
  RGB color space describes colors in terms of the amount of red, green, and blue present. HSV color space describes colors in terms of the Hue, Saturation, and Value. In situations

Fig. 11 **a** Original image and **b** processed image with CIELAB color model

where color description plays an integral role, the HSV color model is often preferred over the RGB model. The HSV model describes colors similarly to how the human eye tends to perceive color. RGB defines color in terms of a combination of primary colors, whereas, HSV describes color using more familiar comparisons such as color, vibrancy and brightness. Transformation from RGB to HSV is as following:

- Hue represents the color type. It can be described in terms of an angle on the above circle. Although a circle contains 360 degrees of rotation, the hue value is normalized to a range from 0 to 255, with 0 being red.
- Saturation represents the vibrancy of the color. Its value ranges from 0 to 255. The lower the saturation value, the more gray is present in the color, causing it to appear faded.
- Value represents the brightness of the color. It ranges from 0 to 255, with 0 being completely dark and 255 being fully bright.
- White has an HSV value of 0–255, 0–255, 255. Black has an HSV value of 0–255, 0–255, 0. The dominant description for black and white is the term, value. The hue and saturation level do not make a difference when value is at max or min intensity level.

- RGB versus YCbCr
  Real-time images and videos are stored in RGB color space, because it is based on the sensitivity of color detection cells in the human visual system. In digital image processing the YCbCr color space is often used in order to take advantage of the lower resolution capability of the human visual system for color with respect to luminosity. Thus, RGB to YCbCr conversion is widely used in image and video processing. Transformation from RGB to *YCbCr* is as following:
  If we have a digital pixel represented in RGB format, 8 bits per sample, where 0 and 255 represents the black and white color, respectively, the YCbCr components can be obtained according to equations.

$$Y = 16 + \left(\frac{65.738R}{256}\right) + \left(\frac{129.057G}{256}\right) + \left(\frac{25.064B}{256}\right)$$

$$Cb = 128 - \left(\frac{37.945R}{256}\right) - \left(\frac{74.494G}{256}\right) + \left(\frac{112.439B}{256}\right)$$

$$Cr = 128 + \left(\frac{112.439R}{256}\right) - \left(\frac{94.154G}{256}\right) - \left(\frac{18.285B}{256}\right)$$

Motion base When use of video sequence is available, motion information can be used to locate moving objects. Moving face and body parts can be extracted by simply thresholding accumulated frame differences (Erik and Low 2001). Besides face regions, facial features can be located by frame differences.

Gray scale base Gray information within a face can also be treated as important features. Facial features such as eyebrows, and lips appear generally darker than their surrounding facial regions (Cootes et al. 1992). Various recent feature extraction algorithms search for local gray minima within segmented facial regions. In these algorithms, the input images are first enhanced by contrast-stretching and gray scale morphological routines to improve the quality of local dark patches and thereby make detection easier. Extraction of dark patches is achieved by low-level gray-scale thresholding based method and consisting of three levels. This system utilizes hierarchical face location consisting of three levels (Erik and Low 2001). Moreover, this algorithm provides an efficient result in complex background where size of the face is unknown.

Edge base Face detection based on edges was introduced (Sakai et al. 1972). This work was based on analyzing line drawings of the faces from photographs, aiming to locate facial features. After that, a hierarchical framework was proposed to trace a human head outline (Craw et al. 1987). A simple and fast system for face detection have been presented (Anila and Devarajan 2010). They proposed framework which consists of three steps i.e. initially the images are enhanced by applying a median filter for noise removal and histogram equalization for contrast adjustment. In the second step the edge image is constructed from the enhanced image by applying Sobel operator. Then a novel edge tracking algorithm is applied to extract the sub windows from the enhanced image based on edges. Further they used Back Propagation Neural Network (BPN) algorithm to classify the sub-window as either face or non-face.

4.1.3 Feature analysis

These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions varies, and then use these to locate faces. These methods are designed mainly for face localization.

Feature searching Viola and Jones presented an approach for object detection which minimizes computation time while achieving high detection accuracy. Viola and Jones proposed a fast and robust method for face detection which is 15 times quicker than existing techniques at the time of release with 95% accuracy (Fig. 12). The technique relies on the use of simple Haar-like features that are evaluated quickly through the use of a new image representation. Based on the concept of an integral image it generates a large set of features and uses the boosting algorithm AdaBoost to reduce the over complete set (Zhang et al. 2011). The detector is applied in a scanning fashion and used on gray-scale images, the scanned window that is applied can also be scaled, as well as the features evaluated. This face detection framework is capable of processing images extremely rapidly while achieving high detection rates. There are three key supports.

- The first one is the introduction of a new image illustration called the integral image which allows the features used by our detector to be computed very quickly.

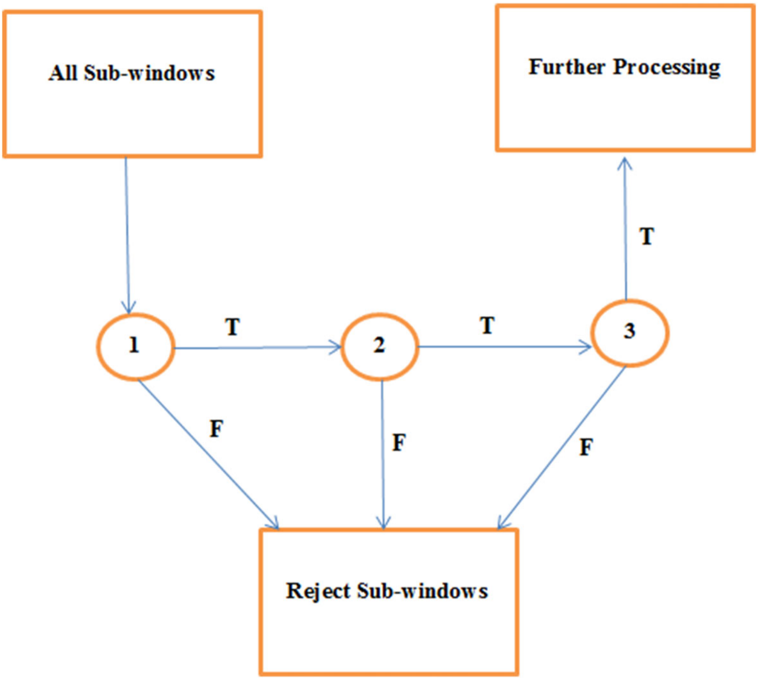


Fig. 12 Working of Viola–Jones methodology

- The second is an easy and efficient classifier which is built using the AdaBoost learning algorithm to select a small number of critical visual features from a very large set of potential features.
- The third contribution is a process for combining classifiers in a cascade which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions.

Advantages:

- It is the most admired algorithm for face detection in real time.
- The main advantage of this approach is uncompetitive detection speed while relatively high detection accuracy, comparable to much slower algorithms.
- Constructing a cascade of classifiers which totally reduces computation time while improving detection accuracy.
- Viola and Jones technique for face detection is an especially successful method as it has a very low false positive rate.

Limitations:

- Extremely long training time.
- Limited head poses.
- Do not detect black Faces.

Local binary pattern (LBP) technique is very effective to describe the image texture features (Ahonen et al. 2004). LBP has advantages such as high-speed computation and rotation invariance, which facilitates the broad usage in the fields of image retrieval, texture examination, face recognition, image segmentation, etc. Recently, LBP was successfully applied to the detection of moving objects via background subtraction. In LBP, every pixel is assigned a texture value, which can be naturally combined with target for tracking thermo graphic and monochromatic video. Major uniform LBP patterns are used to recognize the key points in the target region and then form a mask for joint color-texture feature selection.

Advantages:

- Effective to describe image texture Feature.
- Used in texture analysis, image retrievals, face recognition and image segmentation.
- Detection of moving object via background subtraction.
- Computationally simple than Haar like features and fast.
- The most vital properties of LBP features are tolerance against the monotonic illumination changes and computational simplicity.

Limitations:

- Proposed method is not sensitive to small changes in the face localization.
- Using larger local regions increases the errors.
- It is insufficient for non-monotonic illumination changes.
- Only used for binary and gray images.

AdaBoost algorithm for face detection Boosting is an approach to machine learning based on the idea of creating a highly accurate prediction rule by combining many relatively weak and incorrect rules (Lang and Gu 2009). The AdaBoost algorithm was the first practical boosting algorithm, and one of the most widely used and studied, with applications in numerous field. Using boosting algorithm to train a classifier which is capable of processing images rapidly while having high detection rates. AdaBoost is a learning algorithm which produces a strong classifier by choosing visual features in a family of simple classifiers and combining them linearly. Although AdaBoost is more resistant to over fitting than many machine learning algorithms, it is repeatedly sensitive to noisy data and outliers (Hou and Peng 2009). AdaBoost is called adaptive because it uses multiple iterations to generate a single composite strong learner. AdaBoost creates the strong learner (a classifier that is well-correlated to the true classifier) by iteratively adding weak learners (a classifier that is only slightly correlated to the true classifier).

Filali et al. (2018) have provided a comparative study between four methods (Haar–AdaBoost, LBP–AdaBoost, GF-SVM, GFNN) for face detection. These techniques vary according to the way in which they extract the data and the adopted learning algorithms. The first two methods "Haar–AdaBoost, LBP–AdaBoost" are based on the Boosting algorithm, which is used both for selection and for learning a strong classifier with a cascade classification. While the last two classification methods "GF-SVM, GF-NN" use the Gabor filter to extract the characteristics. Detection time varies from one method to another. Indeed, LBP–AdaBoost and Haar–AdaBoost methods are the fastest compared to others. But in terms of detection rate and false detection rate, the Haar-AdaBoost method remains the best of the four methods. Throughout each round of training, a new weak learner is added to the group and a weighting vector is adjusted to focus on examples that were misclassified in the preceding rounds. The outcome is a classifier that has higher accuracy than the weak learner's classifiers.

Advantages:

- AdaBoost is an algorithm which only needs two inputs: a training dataset and a set of features (classification functions). There is no need to have any prior knowledge about face structure.

- At each stage of the learning, the positive and negative examples are tested by the current classifier. If an example is misclassified, i.e. it cannot be clearly assigned in the good class. In order to increase the discriminant power of the classifier these misclassified examples are up-weighted for the next algorithm iterations.
- The training errors theoretically converge exponentially towards 0. Given a finite set of positive and negative examples, the training error reaches 0 in a finite number of iterations.

Limitations:

- The result depends on the data and weak classifiers. The quality of the final detection depends on the consistency of the training set. Both the size of the sets and the interclass variability are important factors to take into account.
- At each iteration step, the algorithm tests all the features on all the examples which requires a computation time directly proportional to the size of the features and example sets.
- Weak classifiers too complex lead to overfitting.
- Weak classifiers too weak can lead to low margins, and can also lead to overfitting.
- Sensitive to noisy data and outlier.

Gabor features based method An Elastic Bunch Graph Map (EBGM) algorithm that successfully implements face detection system using Gabor filters has been purposed (Sharif et al. 2011). The proposed system applies 40 different Gabor filters on an image. As a result of which 40 images with different angles and orientation are achieved. After that, maximum intensity points in each filtered image are calculated and mark them as fiducial points. The system reduces these points in accordance to distance between them. The next step is calculating the distances between the reduced points using distance formula. At last, the distances are compared with database. If match occurs, it means that the faces in the image are detected.

$$\Psi_{u,v}(z) = \frac{\left\|k_{u,v}\right\|^2}{\sigma^2} e^{\left(\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{i\bar{k}_{u,v}z} - e^{-\frac{\sigma^2}{2}}\right]$$

where

$$\phi_{u,v} = \frac{u\pi}{8}, \quad \phi_u \in [0, \pi]$$

gives the frequency and orientation.

Constellation method All methods discussed so far are able to track faces but still some issues like locating faces of various poses in complex background is truly difficult. To reduce this difficulty scientists, form a group of facial features in face-like constellations using more robust modelling approaches such as statistical analysis. Various types of face constellations have been proposed (Burl et al. 1995). They established use of statistical shape theory on the features detected from a multiscale Gaussian derivative filter. Gaussian filter has been applied for pre-processing in a framework based on image feature analysis (Young and Vliet 1995).

4.2 Image based approaches

4.2.1 Neural network

A rationally attached neural network examines small windows of an image, and chooses whether each window contains a face. The system arbitrates between several networks to enhance performance over a single network. This eliminates the complex task of manually selecting non-face training examples, which must be selected to cover the entire space of non-face images. In early days most hierarchical neural network was proposed (Agui et al. 1992). The first stage is having two parallel subnetworks in which the inputs are filtered the intensity values from an original image. The inputs to the second stage network consist of the outputs of the sub networks and extracted feature values. An output at the second stage shows the presence of a face in the input region. An earliest neural network for face detection is developed, which consists of four layers with 1024 input units, 256 units in the first hidden layer, eight units in the second hidden layer, and two output units (Propp and Samal 1992).

A detection method using auto associative neural networks is presented (Feraud et al. 2001). The idea is based on which shows an auto associative network with five layers is able to perform a nonlinear principal component analysis (Kramer 1991). One auto associative network is used to detect frontal-view faces and another one is used to detect faces turned up to 60 degrees to the left and right of the frontal view. After that a face detection system using Probabilistic Decision-Based Neural Network (PDBNN) has been presented (Lin et al. 1997). The architecture of PDBNN is similar to a radial basis function (RBF) network with modified learning rules and probabilistic interpretation. A multi stage model for face detection is integrated based on Viola and Jones algorithm, Gabor filters, Principal Component Analysis, and Artificial Neural Networks (ANN). The system is composed of two stages: Pre-processing stage and processing stage. A comparison was done between Viola and Jones face detection method (pre-processing stage), and proposed Gabor/PCA and neural network based method (Da'san et al. 2015). Farfade et al. (2015) have proposed a method of face detection based on deep learning, which called Deep Dense Face Detector (DDFD). The method does not require pose/landmark annotation and is able to detect faces in a wide range of orientation using a single model. In addition, DDFD is independent of common modules in recent deep learning object detection methods such as bounding-box regression, SVM, or image segmentation. They compared the proposed method with R-CNN and other face detection methods that are developed especially for multi-view face detection e.g. cascade-based and DPM-based. Liao et al. (2016) proposed a method to address challenges in unconstrained face detection, such as arbitrary pose variations and occlusions. A new image feature called Normalized Pixel Difference (NPD) is proposed. NPD feature is computed as the difference in some ratio between two pixel values, inspired by the Weber Fraction in experimental psychology. The new feature is scale invariant, bounded, and is able to reconstruct the original image. They also proposed a deep quadratic tree to learn the optimal subset of NPD features and their combinations, so that complex face manifolds can be partitioned by the learned rules.

4.2.2 Linear sub-space method

Eigen faces method Eigenvectors has been used in face recognition, in which a simple neural network is demonstrated to perform face recognition for aligned and normalized face images. Images of faces can be linearly encoded using a modest number of basis images (Kirby and Sirovich 1990). They call the set of optimal basis vectors Eigen pictures since these are simply the eigenvectors of the covariance matrix computed from the vectorized face

Fig. 13 Examples of Eigen faces

images in the training set (Hotelling 1933). Experiments on a set of 100 images show that a face image of 91×50 pixels can be effectively encoded using only 50 Eigen faces, while retaining a reasonable likeness (i.e., capturing 95 percent of the variance).

4.2.3 Statistical approach

Support vector machine (SVM) SVMs have also been used for face detection (Mingxing et al. 2013). SVMs work as a new paradigm to train polynomial function, neural networks, or radial basis function (RBF) classifiers. SVMs work on induction principle, called structural risk minimization, which targets 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. Based on two test sets of 10,000,000 test patterns of 19×19 pixels, their system has slightly lower error rates and runs approximately 30 times faster than the system. SVMs have also been used to detect faces and pedestrians in the wavelet domain.

Principal component analysis (PCA) PCA is a technique based on the concept of Eigen faces (Kirby and Sirovich 1990). Turk and Pentland proposed PCA to face recognition and detection. Similarly, PCA on a training set of face images is performed to generate the Eigen faces in face space (Turk and Pentland 1991). Images of faces are projected onto the subspace and clustered. Similarly, non-face training images are projected onto the same subspace and clustered. To detect the presence of a face in a scene, the distance between an image region and the face space is computed for all locations in the image. The result of calculating the distance from face space is a face map (Fig. 13).

4.3 Comparative study of feature based approach and image based approach

In Feature-Based Method 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, 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 skin-color, 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.

In appearance (Image) based methods template 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 non-face images. The learned characteristics are in the form of distribution models or discriminant functions that is consequently used for face detection. Meanwhile, dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy. Its examples are Neural-networks, HMM, SVM, and Adaboost Learning Many appearance-based methods.

	Feature based approach	Image based approach
Technique	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, so there must exist properties or features which are invariant over these variabilities. Facial features such as skin-color, 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	Templates 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 non-face images. The learned characteristics are in the form of distribution models or discriminant functions that is consequently used for face detection
Examples	Skin color, motion, edge, Viola–Jones, snakes etc.	Neural-networks, HMM, SVM, AdaBoost learning etc.
Advantages	Easy to implement	Difficult to implement
Disadvantages	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	Dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy

5 Standard database for face detection

Face image databases are collection of different type of faces, which may be used as test set for face detection system. Some standard face image databases are available, which are following.

Database	Website	Description
MIT dataset	http://cbcl.mit.edu/softwaredatasets/FaceData2.html	19×19 Gray-scale PGM format images Training set: 2429 faces, 4548 non-faces Test set: 472 faces, 23,573 non-faces
PIE database, CMU	www.ri.cmu.edu	A database of 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions
FERET database	www.itl.nist.gov/iad/humanid/feret/feret_master.html	It consists of 14,051 eight-bit gray-scale images of human heads with views ranging from frontal to left and right profiles
The Yale face database	www.face-rec.org/databases/	Contains 165 gray-scale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink
Indian face database	www.pics.stir.ac.uk/Other_face_databases.htm	11 images of each of 39 men, 22 women from Indian Institute of Technology Kanpur
AR database	http://www2.ece.ohio-state.edu/~aleix/	It contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women). Features based on frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf)
SCface—surveillance cameras face database	www.scface.org	Images were taken in uncontrolled indoor environment using five video surveillance cameras of various qualities. Database contains 4160 static images (in visible and infrared spectrum) of 130 subjects

6 Available facial recognition APIs

- *Kairos* Offers a wide variety of image recognition solutions through their API. Their API endpoints include identifying gender, age, emotional depth, facial recognition in both photo and video, and more.
- *Trueface.ai* One flaw with some facial recognition APIs is that they are unable to differentiate between a face and a picture of a face. TrueFace.ai solves that problem with their ability to do spoof detection through their API.
- *Amazon recognition* This facial recognition API is fully integrated into the Amazon Web Service ecosystem. Using this API will make it really easy to build applications that make use of other AWS products.
- *Face recognition and face detection by Lambda Labs* With over 1000 calls per month in the free pricing tier, and only $0.0024 per extra API call, this API is a really affordable option for developers wanting to use a facial recognition API.
- *EmoVu by Eyeris* This API was created by Eyeris and it is a deep learning-based emotion recognition API. EmoVu allows for great emotion recognition results by identifying facial micro-expressions in real-time.
- *Microsoft face API* One cool feature that I found while doing research on the Microsoft Face API, is that the API has the ability to do "similar face search." When this API endpoint is given a collection of faces, and a new face as a query, the API will return a collection of similar faces from the collection.
- *Animetrics face recognition* Using advanced 2D-to-3D algorithms, this API will convert a 2D image into a 3D model. The 3D model will then be used for facial recognition purpose.
- *Face++* This API also has an offline SDK for iOS and Android for you to use. The offline SDK does not provide face recognition, but it can perform face detection, comparing, tracking and landmarks, all while the phone does not have cell service.
- *Google cloud vision* By being integrated into the Google Cloud Platform, this API will be a breeze for you to integrate into applications that are already using other Google Cloud Platform products and services.
- *IBM Watson visual recognition* Whether it is faces, objects, colors, or food, this API lets you identify many different types of classifiers. If the included classifiers aren't enough, then you can train and use your own custom classifiers.

7 Conclusion and future work

In recent years face detection has achieved considerable attention from researchers in biometrics, pattern recognition, and computer vision groups. There is countless security, and forensic applications requiring the use of face recognition technologies. As you can see, face detection system is very important in our day to day life. Among the entire sorts of biometric, face detection and recognition system is the most accurate. In this article, we have presented a survey of face detection techniques. It is exciting to see face detection techniques be increasingly used in real-world applications and products. Applications and challenges of face detection also discussed which motivated us to do research in face detection. The most straightforward future direction is to further improve the face detection in presence of some problems like face occlusion and non-uniform illumination. Current research focuses in field of face detection and recognition is the detection of faces in presence of occlusion and non-uniform illumination. A lot of work has been done in face detection, but not in presence of problem of presence of occlusion and non-uniform illumination. If it happens, it will help a lot to face recognition, face expression recognition etc. Currently many companies providing facial biometric in mobile phone for purpose of access. In future it will be used for payments, security, healthcare, advertising, criminal identification etc.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X (2016) TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX symposium on operating systems design and implementation, pp 265–283

Agui T, Kokubo Y, Nagashashi H, Nagao T (1992) Extraction of face recognition from monochromatic photographs using neural networks. In: Proceeding of 2nd international conference on automation, robotics, and computer vision, vol 1, pp 1881–1885

Ahonen T, Hadid A, Pietikainen M (2004) Face recognition with local binary patterns. In: Proceedings of European conference on computer vision, pp 469–481

Anila S, Devarajan N (2010) Simple and fast face detection system based on edges. Int J Univ Comput Sci 2(1):54–58

Burl MC, Leung TK, Perona P (1995) Face localization via shape statistics. In: Proceedings of international workshop on automatic face and gesture recognition, Zurich, Switzerland, pp 154–159

Cootes TF, Cooper DH, Taylor CJ, Graham J (1992) A trainable method of parametric shape description. In: The proceeding of 2nd British machine vision conference, vol 10, no 5, pp 289–294

Craw I, Ellis H, Lishman JR (1987) Automatic extraction of face-feature. Pattern Recogn Lett 5(2):183–187

Crowley JL, Coutaz J (1997) Vision forman machine interaction. Robot Auton Syst 19(3):347–358

Da'san M, Alqudah A, Debeir O (2015) Face detection using Viola and Jones method and neural networks. In: Proceeding of IEEE international conference on information and communication technology research, pp 40–43

Dhivakar B, Sridevi C, Selvakumar S, Guhan P (2015) Guhan, face detection and recognition using skin color. In: Proceeding of IEEE 3rd international conference on signal processing, communication and networking, pp 1–7

Erik H, Low BK (2001) Face detection: a survey. Comput Vis Image Underst 83:236–274

Farfade SS, Saberian M, Li LJ (2015) Multi-view face detection using deep convolutional neural networks. In: Proceeding of the international conference on multimedia retrieval, pp 1–8

Feraud R, Bernier OJ, Villet JE, Collobert M (2001) A fast and accurate face detector based on neural networks. IEEE Trans Pattern Anal Mach Intell 22(1):42–53

Filali H, Riffi J, Mahraz AM, Tairi H (2018) Multiple face detection based on machine learning. In: Proceeding of international conference on intelligent systems and computer vision, pp 1–8

Hashem HF (2009) Adaptive technique for human face detection using HSV color space and neural networks. In: Proceedings of 26th national radio science conference, pp 1–7

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24(6):417–441

Hou Y, Peng Q (2009) Face detection based on AdaBoost and skin color. In: Proceedings of international symposium on information science and engineering, pp 407–410

Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. In: Proceeding of 1st international conference on computer vision, pp 321–331

Kirby M, Sirovich L (1990) Application of the Karhunen–Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108

Kjeldsen R, Kender J (1996) Finding skin in color images. In: Proceeding of the 2nd international conference on automatic face and gesture recognition, pp 312–317

Kramer MA (1991) Nonlinear principal component analysis using auto associative neural networks. Am Inst Chem Eng J 37(2):233–243

Lang LY, Gu WW (2009) Study on face detection algorithm based on skin color segmentation and AdaBoost algorithm. In: Proceedings of 2nd Pacific-Asia conference on web mining and web-based application, pp 70–73

Liao S, Jain AK, Li SZ (2016) A fast and accurate unconstrained face detector. IEEE Trans Pattern Anal Mach Intell 38(2):211–223

Lin SH, Kung SY, Lin LJ (1997) Face recognition/detection by probabilistic decision-based neural network. IEEE Trans Neural Netw 8(1):114–132

Luo D, Wen G, Li D, Hu Y, Huan E (2018) Deep-learning-based face detection using iterative bounding-box regression. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-5658-5

Mingxing J, Junqiang D, Tao C, Ning Y, Yi J, Zhen Z (2013) An improved detection algorithm of face with combining AdaBoost and SVM. In: Proceeding of 25th Chinese control and decision conference, pp 2459–24633

Mukherjee S, Saha S, Lahiri S, Das A, Bhunia AK, Konwer A, Chakraborty A (2017) Convolutional neural network based Face detection. In: Proceeding of 1st international conference on electronics, materials engineering and nano-technology, pp 1–5

Propp M, Samal A (1992) Artificial neural network architectures for human face detection. In: Proceeding of artificial neural networks in engineering, vol 2, pp 535–540

Ren Z, Yang S, Zou F, Yang F, Luan C, Li K (2017) A face tracking framework based on convolutional neural networks and Kalman filter. In: Proceeding of 8th IEEE international conference on software engineering and service science, pp 410–413

Sakai T, Nagao M, Kanade T (1972) Computer analysis and classification of photographs of human faces. In: Proceeding of 1st USA–Japan computer conference, pp 55–62

Sharif M, Khalid A, Raza M, Mohsin S (2011) Face recognition using Gabor Filters. J Appl Comput Sci Math 11(5):53–57

Subban R, Mishra R (2012) Rule-based face detection in color images using normalized rgb color space—a comparative study. In: Proceeding of international conference on computational intelligence and computing research, pp 1–5

Turk M, Pentland A (1991) Eigen faces for recognition. J Cogn Neurosci 3(1):71–86

Young IT, Vliet LJV (1995) Recursive implementation of the Gaussian filter. Signal Process 44(2):139–151

Yuille AL, Hallinan PW, Cohen DS (1992) Feature extraction from faces using deformable templates. Int J Comput Vis 8:99–111

Zhang H, Xie Y, Xu C (2011) A classifier training method for face detection based on AdaBoost. In: Proceeding of international conference on transportation, mechanical, and electrical engineering, pp 731–734

Zhu X, Ren D, Jing Z, Yan L, Lei S (2012a) Comparative research of the common face detection methods. In: Proceeding of 2nd international conference on computer science and network technology, pp 1528–1533

Zhu Y, Huang C, Chen J (2012b) Face detection method based on multi-feature fusion in YCbCr color space. In: Proceedings of 5th international congress on image and signal processing, pp 1249–1252

Zou L, Kamata S (2010) Face detection in color images based on skin color models. In: Proceeding of IEEE region 10 conferences, pp 681–686