

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Face Recognition: Demystification of Multifarious Aspect in Evaluation Metrics

---

Mala Sundaram and Ambika Mani

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62825>

---

## Abstract

Face recognition has become an interesting research area in the recent era, and blends knowledge from various disciplines such as neuroscience, psychology, statistics, data mining, computer vision, pattern recognition, image processing, and machine learning. A new opportunity is obtained using the application of statistical methods for evaluating the performance of the system. Evaluation methods are the yardstick to examine the efficiency and performance of any face recognition system. Methods for performance evaluation seek to distinguish, compare, and interpret the various factors such as characteristics of subjects, location, illumination, and images. In this chapter, we show how to adapt popular performance measures commonly used in face recognition research, including—precision, recall, *F*-measure, fallout, accuracy, efficiency, sensitivity, specificity, error rate, receiver operating characteristics (ROC). This work serves as an introduction to performance measures, and as a practical guide for using them in research.

**Keywords:** face recognition, feature extraction, face detection, evaluation metrics, biometric

---

## 1. Introduction

The human face plays an interesting role in conveying people's identity in social interaction, biometric systems, law enforcement, security, and surveillance systems [1]. Variety of applications including biometric face recognition technology showed significant attention using the human face as a key to security [2]. As compared with other biometrics systems using fingerprint, iris, and palm print, face recognition has trenchant advantages because of its noncontact process. Face images can be captured from a distance without concerning the person, and the identification process does not require interacting with the person.

Face recognition is one of the major and rapidly thriving fields over the past two decades. This research area straddles researchers from multiple disciplines including data mining, image processing, pattern recognition, neuroscience, psychology, computer vision, and machine learning, etc. The face recognition system can identify one or more individuals from the still images or video by using a stored database of faces [3, 4]. This is a classification problem focusing on automatic face recognition. The main aspect of the face recognition systems is training the system with images from the known persons and classifying the newly coming test images into one of the classes.

Performance evaluation method is the yardstick to analyse the efficiency of any face recognition system. The assessment is essential for understanding the quality of the model or the technique, for refining parameters in the iterative process of learning and for selecting the most adequate model or strategy from a given set of models or techniques [5]. Several criteria are used to evaluate models for different tasks. This chapter goes through general ideas and the techniques used for evaluating the face recognition systems.

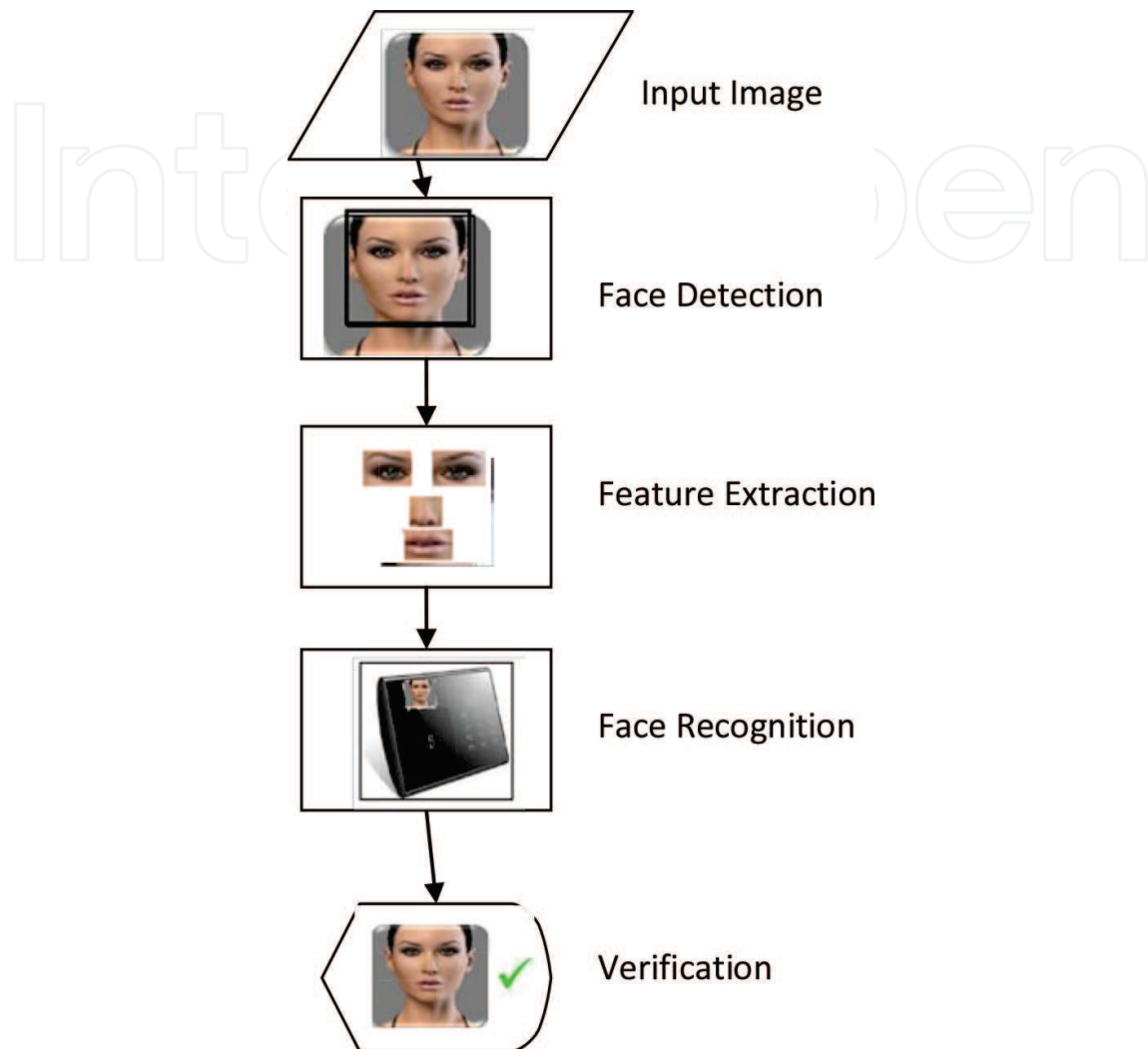
The chapter is structured as follows: Section 2 gives the intricate discussion on the face recognition techniques and methods, Section 3 throws light on the various aspects of the evaluation metrics, Section 4 discuss about the ways of assessing the system, Section 5 details the experimental analysis with case studies, and finally Section 6 concludes the chapter with future direction.

## 2. Face recognition techniques and methods

The human brain is highly adapted for face recognition, by remembering faces better than other patterns, and prefers to look at them over other patterns. Now a days computers also compensates in this research field. Facial recognition systems are applications of computers that examine the digital images of individuals for the purpose of identifying them [6]. The process of face recognition is influenced by many factors such as shape, size, pose, occlusion, and illumination. A human face is an extremely complex object with features that can vary over time. It is covered with nonuniformly textured material skin, which makes face object difficult to model. Skin of the face is influenced by perspiration level. The skin colour changes when the individual is embarrassed or becomes warm.

Facial recognition, have two different applications: basic and advanced. Basic facial recognition identifies faces or nonfaces such as cookies and animals. If it is a face, then the system looks for eyes, a nose, and a mouth. Advanced facial recognition deals with the question on a particular face. This includes unique features: the width of nose, wideness of the eyes, the depth and angle of the jaw, the height of cheekbones, and the distance between the eyes, and creates a unique numerical code. Using these numerical codes, the system then matches that image with another image and identifies how similar the images are to each other. The image sources for facial recognition include pre-existing photos from various databases and video camera signals.

Generally, a face recognition system consists of the following steps: Face detection, feature extraction, and face recognition as in **Figure 1**.



**Figure 1.** General structure of the face recognition system.

### 2.1. Face detection

The main function of this step is to determine the human faces and its location in a given image. The expected outputs are patches within each face or features of the face in the input image. It can also be regarded as object detection to find location and size of all objects in a given image. Face detection could be used for region-of-interest detection, object detection, video and image classification, etc., as in Ref [7–9] (**Figure 2**).

### 2.2. Feature extraction

In this phase, human-face patches are extracted from images to improve the accuracy of face recognition. To recognize human faces, extracting the prominent characteristics on the face

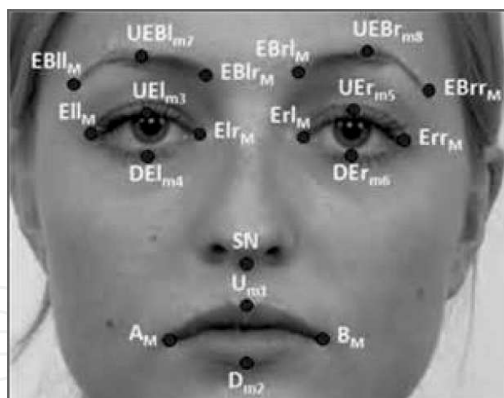


Figure 2. Feature detection.

features such as eyes, nose, and mouth together with their geometry distribution is applied. There are differences in face shape, size, and structure of these organs, so the faces are differing in thousands of ways so as to recognize them. One familiar technique is to extract the shape of the nose, eyes, chin, and mouth, and then distinguish the face by distance and size of those organs. The next method is to use a flexible model to illustrate the shape of the organs on face cleverly. A face patch is next transformed into a feature vector with rigid dimension (Figure 3).

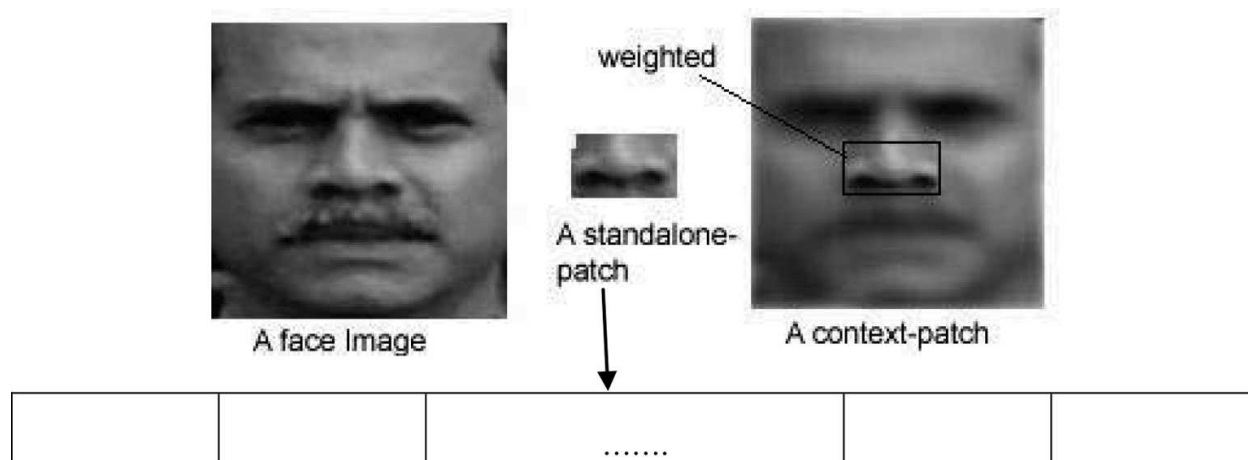


Figure 3. Feature extraction and feature vector representation.

### 2.3. Face recognition

Recognition of face from feature extraction and feature vector representation is the final step. A face data base is needed to achieve an automatic recognition. In the face database, for each person, several images are taken and their characters are stored. When an input face image comes in, the face detection and feature extraction are performed first. Then compare the characteristic features to each face of class stored in the database. The common approach of face recognition is identification and verification [10]. In face identification, the system probes

for the given face image to tell who he/she is, while in face verification, given a face image, the system validates true or false about the identification (Figure 4).

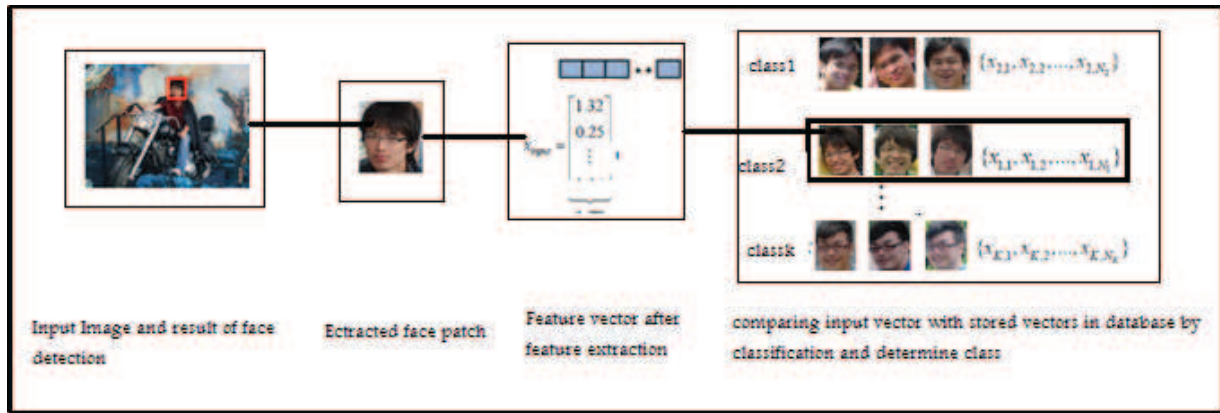


Figure 4. Steps in face recognition system.

### 3. Multifarious aspect in evaluation metrics

Now a days various measures are utilized for evaluating the performance of the face recognition system. This section elaborates some of them. The standard approach to deal with face recognition system evaluation revolves round the ground truth notion of positive and negative detection. Table 1, shows the confusion matrix. The terms positive and negative reveal the asymmetric condition on detection tasks where one class is the relevant pattern class and another class is the nonrelevant class.

Ground truth\detection	Detected (Positive)	Rejected (Negative)
Relevant	True positive (TP)	False negative (FN)
Nonrelevant	False positive (FP)	True negative (TN)

Table 1. Confusion matrix.

In the case of binary recognition or two class recognition, the system has to differentiate between face and nonface criteria. The true positive means the portion of face images to be detected by the system, while the false positive means the portion of nonface images to be detected as faces. The term true positive here has the same meaning as the detection rate and recall. False positives implies wrongly matching the individuals with photos in the database, and false negatives means not catching people even when their photo is in the database. There are two main evaluation plots: the receiver operating characteristics (ROC) curve and the precision and recall (PR) curve. The ROC curve examines the relation between the true positive rate and the false positive rate, while the PR curve extracts the relation between detection rate (recall) and the detection precision.

### 3.1. Precision

Precision is the fraction of the detected images that square measure relevant to the user's wants. It is additionally referred to as reliability or repeatability and is that the degree to that recurrent measurements beneath unchanged conditions show an equivalent results. Equation (1) represents them.

$$\text{Precision} = \frac{\text{No of true positive}}{\text{No of all detected patterns}} \quad (1)$$

In binary classification, precision is additionally known as positive predictive value. It is represented in Equation (2).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

### 3.2. Recall

Recall is the proportion of positive cases that were properly identified. It is the fraction of relevant images that are successfully detected. It is additionally referred to as true positive rate. Recall is calculated using Equation (3).

$$\text{Recall} = \frac{\text{No of true positive}}{\text{No of relevant patterns}} \quad (3)$$

In binary classification, recall is commonly referred to as sensitivity. It is denoted in Equation (4).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

### 3.3. Fall out

Fall out is the proportion of nonrelevant images that are detected as positive, out of all nonrelevant images (Equation 5).

$$\text{Fallout} = \frac{|\{non-relevant\} \cap \{detected\}|}{|\{non-relevant\}|} \quad (5)$$

In case of binary category, fallout is closely associated with specificity and is capable (1 – specificity). It is often checked out as the chance that nonrelevant images are detected as positive (Equation 6).

$$Fallout = \frac{TN}{TN + FP} \quad (6)$$

### 3.4. F-measure

F-measure is additionally referred to as F-Score or F1-measure. It combines the exactness and recall. It computes the average of the precision and recall. A conventional F-measure is the harmonic mean of precision and recall. This score is used to give a summary of the PR curve. It will be denoted as in Equation 7:

$$F\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

In binary classification it is denoted as in Equation 8:

$$F\text{-measure} = \frac{2 \times TP}{(2 \times TP + FP + FN)} \quad (8)$$

The harmonic mean is an additional intuitive then the arithmetic mean, once computing the quantitative relation. Therefore, the complete definition of F-measure is given by Equation 9.

$$F\text{-measure} = \frac{(\beta^2 + 1) PR}{\beta^2 P + R} \quad (9)$$

$$\text{where } \beta^2 = \frac{1-\alpha}{\alpha} \quad \text{and} \quad \alpha \in [0,1] \text{ and } \beta \in [0,\infty]$$

$$\alpha = \frac{1}{2} \text{ or } \beta = 1 \text{ is commonly written as } F_1 \text{ or } F_{\beta=1} = \frac{2PR}{P + R}$$

$\beta$  is the parameter that controls a balance between  $P$  and  $R$ . When  $\beta = 1$ ,  $F_1$  involves be similar to the harmonic mean of  $P$  and  $R$ . This is often also referred to as F-measure or balanced F-score since precision and recall are equally weighted. When  $\beta > 1$  emphasize recall. When  $\beta < 1$  emphasize precision.

### 3.5. Accuracy

Accuracy is the proportion of classifications, over all the  $N$  examples that were correctly detected. Accuracy is defined as “the fraction of quantity of correct classification over the entire number of samples.” The amount of predictions in classification techniques relies upon the counts of the test records properly or incorrectly predicted by the model [11]. These counts are tabulated into a confusion matrix (also referred as contingency) Table 1. The confusion matrix shows how the classifier is behaving for individual categories.



$$\text{Accuracy} = \frac{\text{No of correctly detected pattern}}{\text{Total number of validation set}} \quad (10)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

### 3.6. Error rate

The fraction is the range quantity of misclassification over the overall number of validation samples. However, the system response to wrong answers is the motive behind the introduction of error rate. It is an acceptable performance measure for the comparison of classification techniques given the balanced datasets. Precision, recall, and  $F$ -measure are acceptable performance measures for unbalanced datasets (Equations 13 and 14).

$$\text{Error rate} = \frac{\text{No of misclassification}}{\text{No of samples in the validation set}} \quad (12)$$

$$e = \frac{FP+FN}{TP+TN+FP+FN} \quad (13)$$

### 3.7. Effectiveness

The effectiveness measure is based on  $F_\beta$ -measure.  $F_\beta$  "Measures the effectiveness of detection with respect to a user who attaches  $\beta$  times as much importance to recall as precision (Equation 14)."

$$E \text{ (effectiveness) is } E(P, R) = 1 - \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (14)$$

where  $\beta$  determines the relative importance of precision ( $P$ ) and recall ( $R$ )

### 3.8. Sensitivity

True positive rate (TPR) is named sensitivity, hit rate, and recall. An applied mathematical measure of how well a binary classification test properly identifies a condition probability of properly labelling members of the target class (Equation 15).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (15)$$

### 3.9. Specificity

True negative rate (TNR) is named specificity. It is an applied mathematics measure of how well a binary classification test properly identifies the negative cases (Eq. 16).

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (16)$$

False positive rate (FPR) also called as alarm rate is denoted as in Eqs. 17 and 18:

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (17)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{False alarm} \quad (18)$$

### 3.10. Receiver operating characteristics

Receiver operating characteristics (ROC) is a graph used for organizing and visualizing the performance of a system. It is a distinct option for precision–recall curves [12]. ROC graphs are normally utilized in medical decision-making, and in recent years are used more and more in machine learning and data processing research. It is a graphical representation for displaying the transition between TPR and FPR. TPR indicates correctly classified or total positive values and plotted on the  $y$ -axis, whereas FPR indicates incorrectly classified or total negative values plotted on the  $x$ -axis.

The points on the top left of ROC have high TP Rate and low FP Rate, thus represents smart classifiers. ROC graphs are far more helpful for domains with skew category distribution and unequal classification error costs. For this ability, ROC graphs are far more popular than accuracy and error rate. ROC plot can also visualize characterization change between the False match rate (FMR) and False nonmatch rate (FNMR).

Generally, the matching technique performs a decision based on a threshold that determines how close the image is to a template. If the threshold is reduced, there will be fewer false nonmatches, but more false accepts. Similarly, a higher threshold will reduce the FMR, but increase the FNMR. This more linear graph illuminates the differences for higher performances (rarer errors).

In **Figure 5**, the value A depicts Conservative performance which makes positive performance only with a strong evidence, so few false positive errors. The value B indicates the Liberal performance and value C indicates the perfect performance.



Figure 5. Regions of ROC graphs.

Some of the additional measures to evaluate the performance of Face identification systems are the following: Recognition Rate, Verification Rate, Half Total Error Rate in Ref. [13], Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), and False Rejection Rate (FRR)

The Recognition Rate is the simplest measure. It relies on a list of gallery images (usually one per identity) and a list of probe images of the same identities. The Recognition Rate is the total number of correctly identified probe images divided by the total number of probe images.

Another evaluation measure is the Verification Rate as in Ref. [14]. It relies on a list of image pairs, where pair with the same and pairs with different identities are compared. Given the lists of similarities of types, the ROC graph can be computed, and finally the Verification Rate. There are some more measures, such as the Half Total Error Rate and similar, which rely on independent development and evaluation sets. Validation test is a kind of test used to identify faces. The verification system uses some measures (i.e., Equal Error Rate), while some other are usually adopted for recognition systems (i.e., Recognition Rate).

### 3.11. False match rate

It is also denoted as FMR or False Accept Rate (FAR). FMR is the probability that the system incorrectly matches the input pattern to a nonmatching template in the database. It gauges the percent of invalid inputs that are incorrectly accepted. Similarly, if the person is an imposter in reality, but the matching score is higher than the threshold, then he is treated as genuine. This increases the FMR also depends upon the threshold value.

### 3.12. False nonmatch rate

It is also denoted as FNMR or false reject rate (FAR). It is the probability that the system fails to detect a match between the input pattern and a matching template from the database. It measures the percent of valid inputs that are incorrectly rejected.

### 3.13. Equal error rate

It is denoted as crossover error rate (EER or CER) or the rate at which both acceptance and rejection error are equal. The value of the EER can be obtained from the ROC curve. The EER is a quick way to compare the accuracy of devices with different ROC curves. Normally, the device with the lowest EER is the most accurate.

### 3.14. Failure to enroll rate

Also represented as FTE or FER is the rate at which endeavors to create a template from an input is unsuccessful. This case is usually caused by low-quality inputs.

### 3.15. Failure to capture rate

FTC is the probability that the system fails to detect an input even when the input is presented correctly.

## 4. Evaluation of face recognition system

Recognition of faces relies on how flexible the system is for pose variations. If the aim of the system is to recognize only frontal faces, then just use few classifier and function. The number of images of each face relies on the training on an image and testing on the rest. The recognition from different angles depends on the type of images and training set accordingly, with at least one image for each pose per person. The number of images in the training set and test on the remainder depends on the application of the system.

There are three methods to measure accuracy in a face recognition task. The one that was most suitable might depend to an extent on what the end purpose was.

1. How accurate is the algorithm at detecting a person from a data set containing *many images of a person and various images of different people*.
2. How accurate is the algorithm at gaining knowledge with a *set of faces from training and testing datasets of same peoples' images*.
3. How accurate is the algorithm at identifying *more than one person* from a dataset containing *images of these people mixed with the other people*.

For Case 1, train the algorithm with a set of images of an individual person's face and test on a set of images that contain different images of the goal person as well as equal number of other people. This task would be a binary classification task and accuracy can be efficiently measured with the help of precision and recall then. For more generalized results, this test could be repeated using various people.

For Case 2, train on multiple images of several people and then test on different images of the same people (If the dataset contains limited persons, then leave-one-out methodology might

be useful). This type of multiclass classification problem can be evaluated with the help of confusion matrixes which would be helpful in evaluating this sort of test.

For Case 3, train the algorithm on a categorized training set of images of several people and then test on a set of images containing different images of the same people mixed with other images of faces (To recognize people from a crowd, then large number of different peoples' images can be mixed in the test dataset). This could be created as a binary classification (person of interest/not), or as a multiclass problem (each person is a separate class with others). If the test set contains unbalanced images, then various measures of accuracy with true negatives can be used.

## 5. Experimental analysis

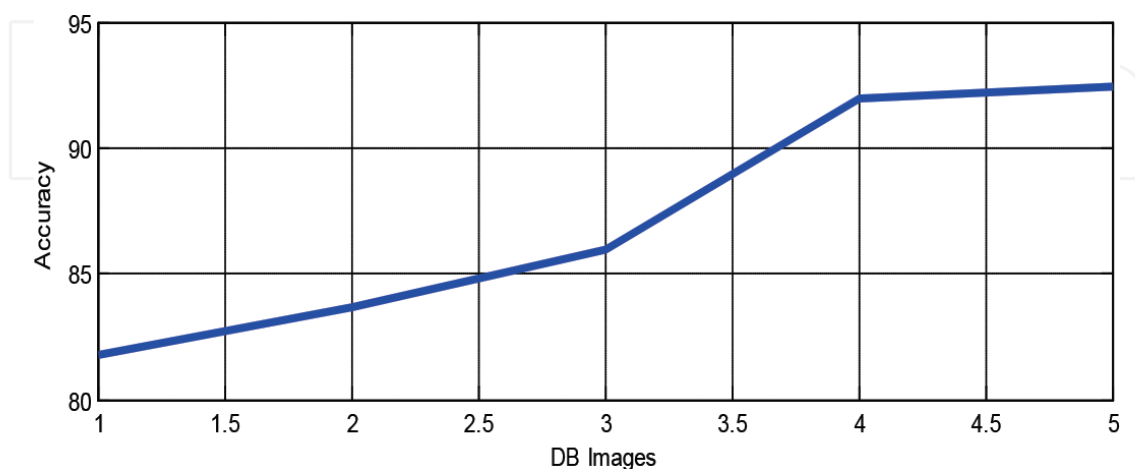
Face recognition has various result challenges as in reference [15]. In this section we have employed the theoretical model for computing the various performance measures to evaluate the efficiency of the face recognition system in different aspects.

### 5.1. Case 1

This case study used publicly available AT&T database in reference [16] for recognition experiments. In the database, 10 different images of each of 40 persons (total 400 images) with deviations in angles, expressions, and facial details are conceived. A preview image of the Database of Faces is shown in **Figure 6**.

The comparison is performed using Support Vector Machine technique and the computational efficiency is tabulated in the Table 2 and depicted in **Figure 7**.

**Figure 8** shows the accuracy measure of the various datasets obtained using various technique.



**Figure 6.** Sample collection of images in the dataset.

S.No	DB Images	Accuracy (Recognition Rate)
1	3	81.80%
2	4	83.70%
3	5	86.00%
4	6	92.00%
5	7	92.40%

Table 2. Accuracy of the recognition system.

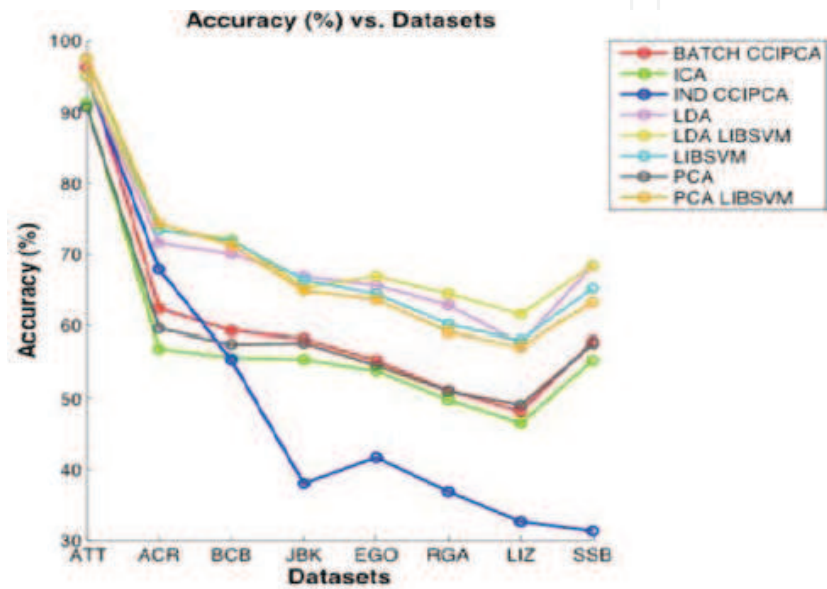


Figure 7. Accuracy of the recognition system.

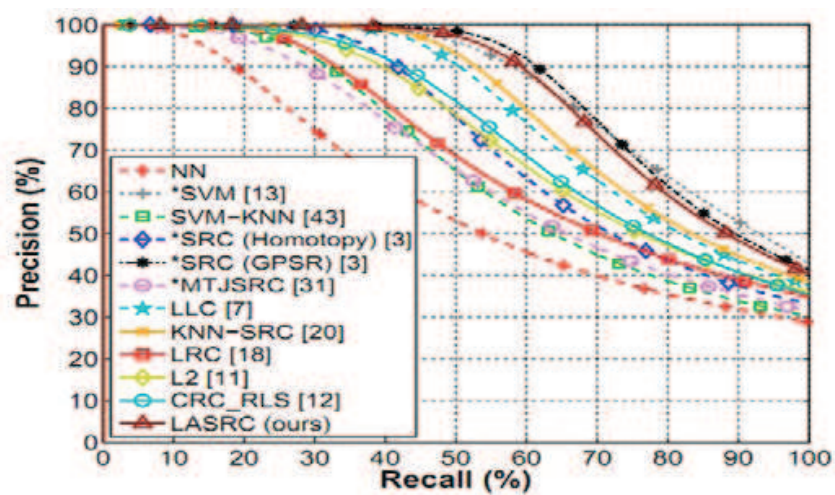
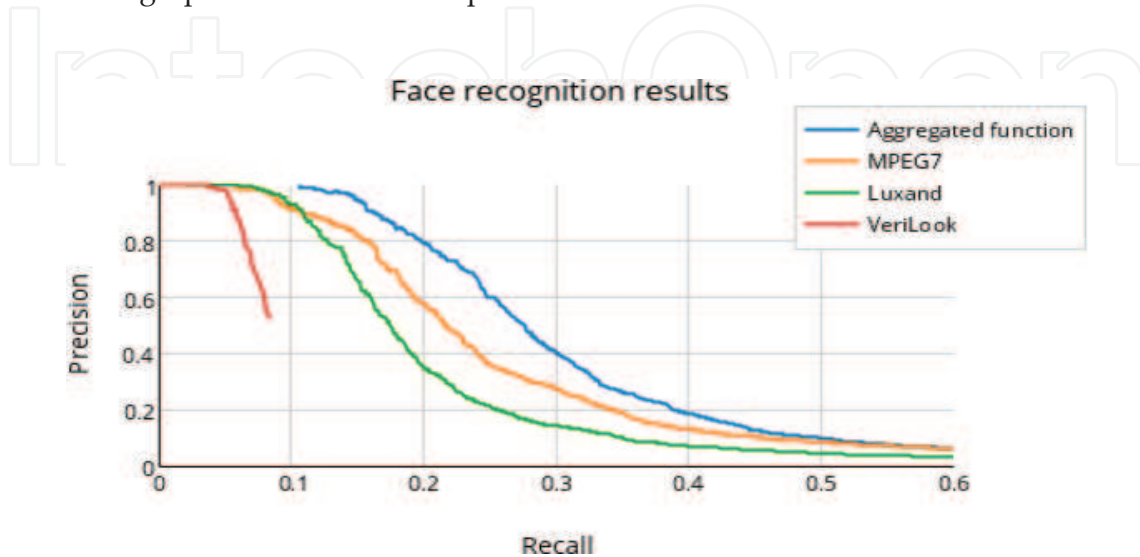


Figure 8. Accuracy of the recognition system using various datasets.

## 5.2. Case 2

Brian C. Becker, gathered 800,000 face dataset from the Facebook social network as in reference [17] that models real-world situations where specific faces must be recognized and unknown identities must be rejected. Finally, the results are depicted using precision–recall curve as in **Figure 9**. The graph shows that as the precision increases recall decreases.



**Figure 9.** Precision and recall curve on our 800,000 Facebook dataset.

Nonreal time algorithms are marked with an asterisk (\*). LASRC approach performs very similarly to nonreal time algorithms such SRC or SVMs but has the advantage of being real time. In fact, LASRC trains 100× faster than SVMs and classify 250× faster than SRC. Compared to other real-time methods, LASRC outperforms state-of-the-art least squares, sparse, and max-margin classifiers.

Face recognition is a technology for automatic detection and recognition of human faces on static images as stated in reference [18]. The main advantage of this technology is its ability to aggregate multiple face recognition and detection functions. Here we listed some of the commercial software for face recognition such as FaceSDK, VeriLook SDK, MPEG-7 descriptors + OpenCV. The following Table 3 and **Figure 10** show the values of precision and recall obtained using the listed software.

Name	Recall	Precision
OpenCV	55%	89%
FaceSDK	63%	83%
VeriLook SDK	73%	84%
Aggregation approach	62%	98%

**Table 3.** Comparison results of precision and recall.

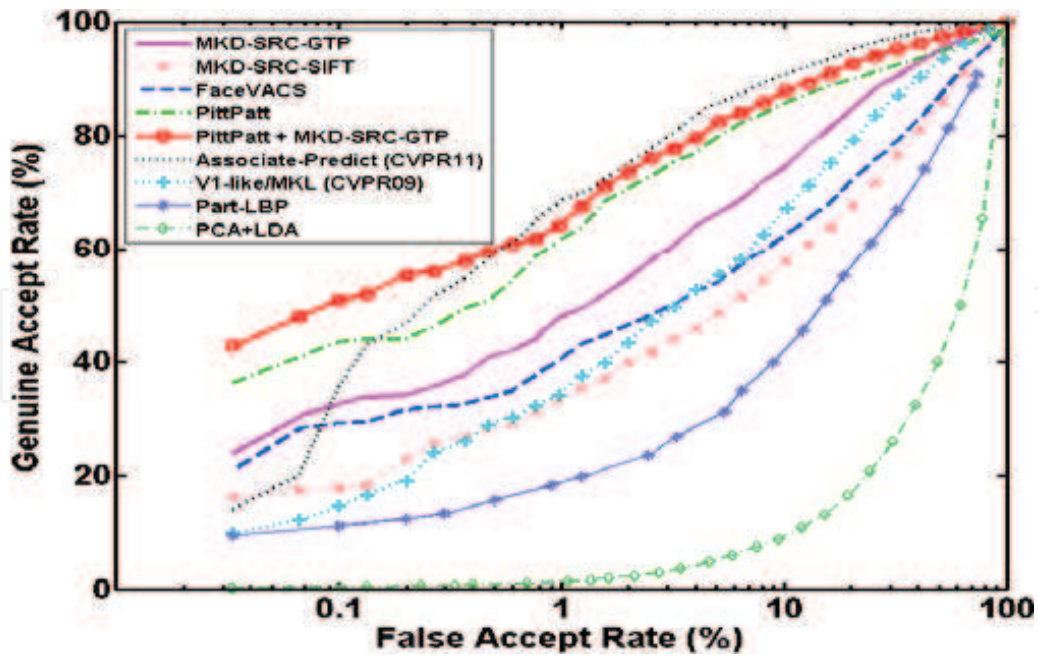


Figure 10. Comparison results of precision and recall.

### 5.3. Case 3

This case study used the LFW benchmark dataset, where the dataset is divided into 10 subsets for cross validation, with each subset containing 300 pairs of genuine matches and 300 pairs of impostor matches for verification. The mean values of FAR and Genuine Accept Rate (GAR) with fixed thresholds over all the 10 subsets are plotted in an ROC curve for **performance evaluation** as in reference [19] and **Figure 11**.

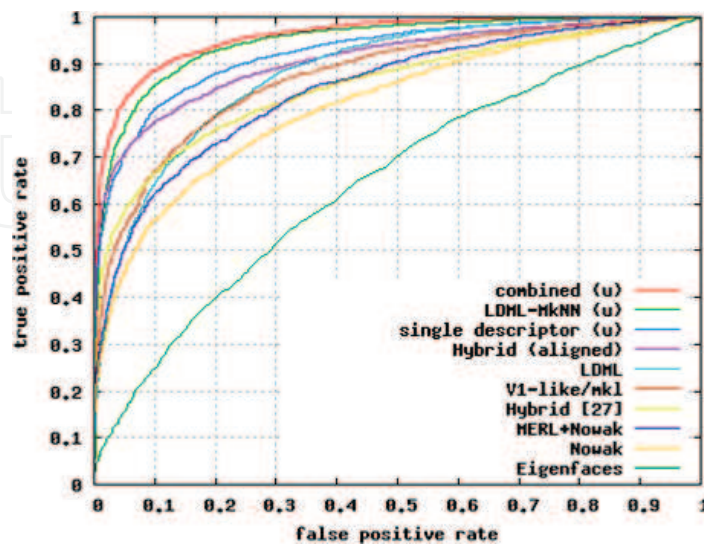


Figure 11. The ROC curves of the various face recognition algorithms.



The following ROC curves (**Figure 12**) are the average over ten-folds (FPR and TPR) of the LFW data set. The (u), indicates ROC curve is for the unrestricted setting.

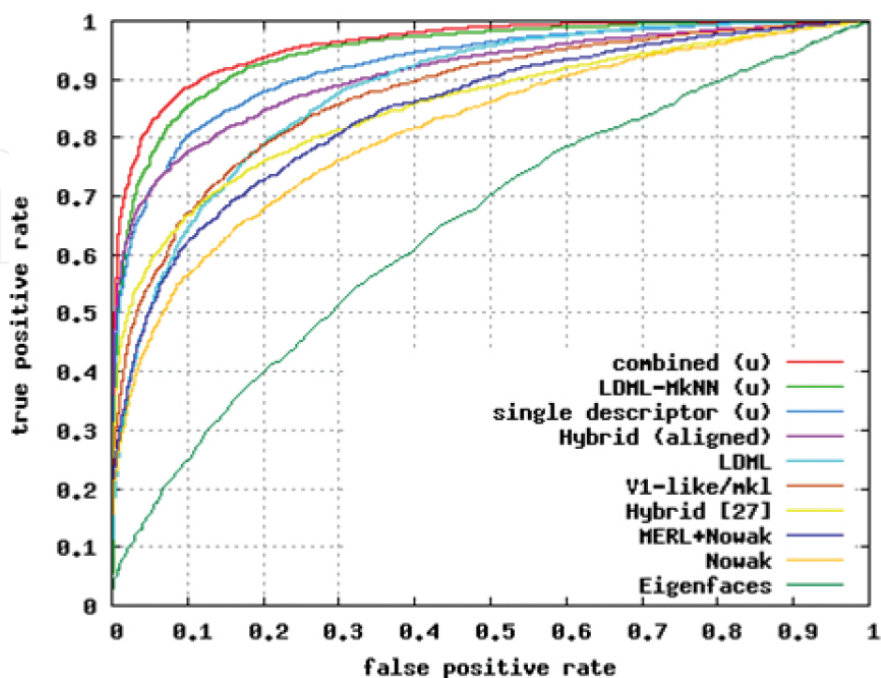


Figure 12. The ROC curves using TPR and FPR.

## 6. Conclusion and future work

This chapter presents a viewpoint about face recognition and the various ways to evaluate the face recognition system. The faces are highly complex patterns that often differ in only subtle ways, like changes in angle and lighting. Hence, the face recognition system should consider various factors such as facial expression change, aging, pose change, illumination change, scaling factor, frontal vs. profile presence and absence of spectacles, occlusion due to scarf, mask in front, beard, and moustache. Generally, when the training set contains faces of one person, then precision and recall could be used to evaluate accuracy. When the training set contains multiple faces of several people and test set contains the different faces of same people, then confusion matrixes would be helpful in evaluating the test face. When the training contains faces of interest with other faces, and the test set is an unbalanced one, then various measures of accuracy dominated by true negatives can be used to evaluate the face recognition. A complete face recognition system contains several subproblems where each one is an independent research problem. The line of future work includes the assessment of various machine learning algorithms used in face recognition with feature mining. However, next era face recognition are going to have tremendous application in smart environs, real time, and in much less-controlled situations.

## Author details

Mala Sundaram\* and Ambika Mani

\*Address all correspondence to: [ursmala@gmail.com](mailto:ursmala@gmail.com)

Department of Computer Science and Engineering, Anna University (BIT Campus), Tamil Nadu, India

## References

- [1] Kumar A., Kumar A. Adaptive security for human surveillance using multimodal open Set biometric recognition. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR); 2014. pp. 405–410. DOI: 10.1109/ICPR.2014.78
- [2] Hitoshi I. The world's best face recognition system to achieve safety and security in our society. Special Issue on Solutions for Society—Creating a Safer and More Secure Society NEC Technical Journal 2015. Vol.9, No.1, 48–51.
- [3] Liao S., Jain A., Li S. Partial face recognition: alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 2013. Vol. 35, Issue 5, 1193–1205. DOI: 10.1109/TPAMI.2012.191
- [4] Mohanty P., Sarkar S., Kasturi R., Phillips P. Subspace approximation of face recognition algorithms: an empirical study. IEEE Transactions on Information Forensics and Security 2008. Vol. 3, Issue 4, 734–748. DOI: 10.1109/TIFS.2008.2007242
- [5] W-L. Chao. Integrated machine learning algorithms for human age estimation [thesis]. National Taiwan University; 2011.
- [6] Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face recognition: a literature survey. ACM Computing Surveys (CSUR) 2003. Vol. 35, Issue 4; 399–458. DOI: 10.1145/954339.954342
- [7] Low E.H. Face detection: a survey. Computer Vision and Image Understanding 2001. Vol. 83, 236–274. DOI: 10.1006/cviu.2001.0921
- [8] Ming-Hsuan Y., D.J. Kriegman, N. Ahuja. Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002. Vol. 24, Issue 1, 34–58. DOI: 10.1109/34.982883
- [9] S. Zafeiriou, C. Zhang, Zhengyou Z..A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding 2015. Vol.138, Issue C, 1–24. DOI: 10.1016/j.cviu.2015.03.015

- [10] Kumar N., Berg A., Belhumeur P., Nayar S. Describable Visual Attributes for Face Verification and Image Search, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2011. Vol. 33, Issue 10, 1962–1977. DOI: 10.1109/TPAMI.2011.48
- [11] Aggarwal G., Biswas S., Flynn P.J., Bowyer K.W. Predicting performance of face recognition systems: An image characterization approach. In: *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW); Workshops, IEEE Press; 2011. pp. 52–59. DOI:10.1109/CVPRW.2011.5981784*
- [12] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* 2006. Vol. 861–874. DOI: 10.1016/j.patrec.2005.10.010
- [13] White D., Dunn J.D., Schmid A.C., Kemp R.I.. Error rates in users of automatic face recognition software. *PLoS ONE* 2015. Vol. 10, Issue 10, e0139827. doi:10.1371/journal.pone.0139827
- [14] P. Ahlgren, L. Grönqvist. Retrieval evaluation with incomplete relevance data: a comparative study of three measures. In: *Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM '06); New York, 2006. pp. 872–873 DOI: 10.1145/1183614.1183773*
- [15] Phillips P.J. Preliminary face recognition grand challenge results. In: *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06), Washington, USA IEEE 2006. pp. 15–24. 10.1109/FGR.2006.87*
- [16] Database of Faces by AT & T [Internet]. 2016. Available from: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> [Accessed: 2016-11-12]
- [17] Brian C. Becker [Internet]. 2016. Available from: <http://www.brianbecker.com/blog/research/web-scale-face-recognition/> [Accessed: 2016-01-27]
- [18] Laboratory of Data Intensive Systems and Applications (DISA)[Internet]. Available from <http://disa.fi.muni.cz/> [Accessed: 2016-01-11]
- [19] Ho, W.H., Watters, P. A new performance evaluation method for face identification— regression analysis of misidentification risk. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007); 2007; pp. 1–6. DOI: 10.1109/CVPR.2007.383276*