Facial Detection and Recognition: A Detailed Survey

[1]Muskan Gupta, [2]Sanjay Singh,
[1]Student, [2]Assistant Professor
[1,2]Department of Computer Science,
[1]Aravali College of Engineering & Management, Faridabad, India

Abstract : Face recognition has been one of the most interesting and important research fields in the past two decades. The reasons come from the need of automatic recognitions and surveillance systems, the interest in human visual system on face recognition, and the design of human-computer interface, etc. These researches involve knowledge and researchers from disciplines such as neuroscience, psychology, computer vision, pattern recognition, image processing, and machine learning, etc. A bunch of papers have been published to overcome difference factors (such as illumination, expression, scale, pose, ……) and achieve better recognition rate, while there is still no robust technique against uncontrolled practical cases which may involve kinds of factors simultaneously. In this report, we'll go through general ideas and structures of recognition, important issues and factors of human faces, critical techniques and algorithms, and finally give a conclusion.

I. INTRODUCTION

In this report, we focus on image-based face recognition. Given a picture taken from a digital camera, we'd like to know if there is any person inside, where his/her face locates at, and who he/she is. Towards this goal, we generally separate the face recognition procedure into three steps: Face Detection, Feature Extraction, and Face Recognition (shown at fig.1 below).

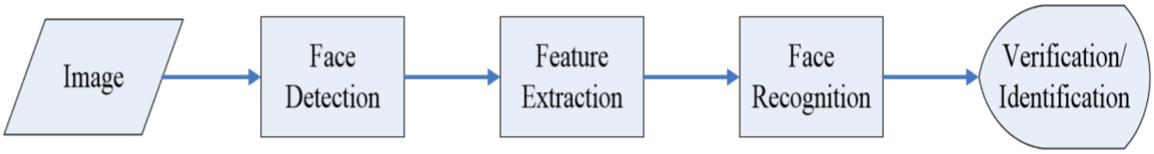

Figure1: Configuration of a general face recognition structure

Face Detection:
The main function of this step is to determine (i) whether human faces appear in a given image, and (ii) where these faces are located at. The expected outputs of this step are patches containing each face in the input image. In order to make further face recognition system more robust and easier to design, face alignment is performed to justify the scales and orientations of these patches. Besides serving as the pre-processing for face recognition, face detection could be used for region-of-interest detection, retargeting, video and image classification, etc.

Feature Extraction:
After the face detection step, human-face patches are extracted from images. Directly using these patches for face recognition have some disadvantages, first, each patch usually contains over 1000 pixels, which are too large to build a robust recognition system1. Second, face patches may be taken from different camera alignments, with different face expressions, illuminations, and may suffer from occlusion and clutter. To overcome these drawbacks, feature extractions are performed to do information packing, dimension reduction, salience extraction, and noise cleaning. After this step, a face patch is usually transformed into a vector with fixed dimension or a set of fiducial points and their corresponding locations. In some literatures, feature extraction is either included in face detection or face recognition.

Face Recognition:
After formulizing the representation of each face, the last step is to recognize the identities of these faces. In order to achieve automatic recognition, a face database is required to build. For each person, several images are taken and their features are extracted and stored in the database. Then when an input face image comes in, we perform face detection and feature extraction, and compare its feature to each face class stored in the database. There have been many researches and algorithms proposed to deal with this classification problem, and we'll discuss them in later sections. There are two general applications of face recognition, one is called identification and another one is called verification. Face identification means given a face image, we want the system to tell who he / she is or the most probable identification; while in face verification, given a face image and a guess of the identification, we want the system to tell true or false about the guess.

II. FUNDAMENTAL OF PATTERN RECOGNITION

Before going into details of techniques and algorithms of face recognition, we'd like to make a digression here to talk about pattern recognition. The discipline, pattern recognition, includes all cases of recognition tasks such as speech recognition, object recognition, data analysis, and face recognition, etc. In this section, we won't discuss those specific applications, but introduce the basic structure, general ideas and general concepts behind them. The general structure of pattern recognition is shown in figure below. In order to generate a system for recognition, we always need data sets for building categories and compare similarities between the test data and each category. A test data is usually called a "query" in image retrieval literatures, and we will use this term throughout this report. From figure, we can easily notice the symmetric structure. Starting from the data sets side, we first perform dimension reduction2 on the stored raw data. The methods of dimension reduction can be categorized into data-driven methods and domain-knowledge methods, which will be discussed later. After dimension reduction, each raw data in the data sets is transformed into a set of features, and the classifier is mainly trained on these feature representations. When a query comes in, we perform the same dimension reduction procedure on it and enter its features into the trained classifier. The output of the classifier will be the optimal class (sometimes with the classification accuracy) label or a rejection note (return to manual classification).

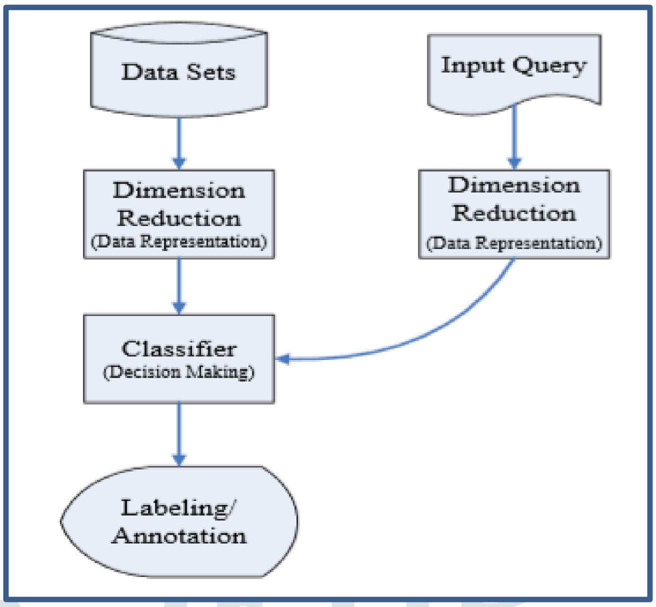

Figure2: General structure of a pattern recognition system

III. Issues and Factors of Human Faces

When focusing on a specific application, besides building the general structure of pattern recognition system, we also need to consider the intrinsic properties of the domain-specific data. For example, to analyze music or speech, we may first transform the input signal into frequency domain or MFCC (Mel-frequency cepstral coefficients) because features represented in these domains have been proved to better capture human auditory perception. In this section, we'll talk about the domain-knowledge of human faces, factors that result in face-appearance variations in images, and finally list important issues to be considered when designing a face recognition system.

Aspects from psychophysics and neuroscience

There are several researches in psychophysics and neuroscience studying about how we human performs recognition processes, and many of them have direct relevance to engineers interested in designing algorithms or systems for machine recognition of faces. In this subsection, we briefly review several interesting aspects. The first argument in these disciplines is that whether face recognition a dedicated process against other object recognition tasks. Evidences that faces are more easily remembered by humans than other objects when presented in an upright orientation and prosopagnosia patients can recognize faces from other objects but have difficulty in identifying the face support the viewpoint of face recognition as a dedicated process. While recently, some findings in human neuropsychology and neuroimaging suggest that face recognition may not be unique.

Holistic-based or feature-based

This is another interesting argument in psychophysics / neuroscience as well as in algorithm design. The holistic-based viewpoint claims that human recognize faces by the global appearances, while the feature-based viewpoint believes that important features such as eyes, noses, and mouths play dominant roles in identifying and remembering a person.

Thatcher Illusion

The Thatcher illusion is an excellent example showing how the face alignment affects human recognition of faces. In the illusion shown in the figure below, eyes and mouth of an expressing face are excised and inverted, and the result looks grotesque in an upright face. However, when shown inverted, the face looks fairly normal in appearance, and the inversion of the internal features is not readily noticed.

Factors of human appearance variations

There are several factors that result in difficulties of face detection and face recognition. Except the possible low quality driven from the image acquisition system, we focus on the angle of human faces taken by the camera and the environment of photo acquisition. There are generally six factors we need to concern: (i) illumination, (ii) face pose, (iii) face expression, (iv) RST (rotation, scale, and translation) variation, (v) clutter background, and (vi) occlusion.

Table 1 lists the details of each factor.

Illumination	The illumination variation has been widely discussed in many face detection and recognition researches. This variation is caused by various lighting environments and is mentioned to have larger appearance difference than the difference caused by different identities. The example of illumination changes on images of the same person, and it's obviously that under some illumination conditions, we can neither assure the identification nor accurately point out the positions of facial features.
Pose	The pose variation results from different angles and locations during the image acquisition process. This variation changes the spatial relations among facial features and causes serious distortion on the traditional appearance-based face recognition algorithms such as Eigen faces and fisher faces.
Expression	Human uses different facial expressions to express their feelings or tempers. The expression variation results in not only the spatial relation change, but also the facial-feature shape change.
RST variation	The RST (rotation, scaling, and translation) variation is also caused by the variation in image acquisition process. It results in difficulties both in face detection and recognition, and may require exhaustive searching in the detection process over all possible RST parameters.
Cluttering	In addition to the above four variations which result in changes in facial appearances, we also need to consider the influence of environments and backgrounds around people in images. The cluttering background affects the accuracy of face detection, and face patches including this background also diminish the performance of face recognition algorithms.
Occlusion	The occlusion is possibly the most difficult problem in face recognition and face detection. It means that some parts of human faces are unobserved, especially the facial features.

Table 1: The list and description of the six general factors

IV. FACE DETECTION

Knowledge-based methods

These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features. These methods are designed mainly for face localization, which aims to determine the image position of a single face. In this subsection, we introduce two examples based on hierarchical knowledge-based method and vertical / horizontal projection.

Hierarchical knowledge-based method

This method is composed of the multi-resolution hierarchy of images and specific rules defined at each image level [4]. The hierarchy is built by image sub-sampling. The face detection procedure starts from the highest layer in the hierarchy (with the lowest resolution) and extracts possible face candidates based on the general look of faces. Then the middle and bottom layers carry rule of more details such as the alignment of facial features and verify each face candidate. This method suffers from many factors described in Section 3 especially the RST variation and doesn't achieve high detection rate (50 true positives in 60 test images), while the coarse-to-fine strategy does reduce the required computation and is widely adopted by later algorithms.

Horizontal / vertical projection

This method uses the fairly simple image processing technique, the horizontal and vertical projection [6]. Based on the observations that human eyes and mouths have lower intensity than other parts of faces, these two projections are performed on the test image and local minimums are detected as facial feature candidates which together constitute a face candidate. Finally, each face candidate is validated by further detection rules such as eyebrow and nostrils.

Feature invariant approaches

These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate faces. These methods are designed mainly for face localization. To distinguish from the knowledge-based methods, the feature invariant approaches start at feature extraction process and face candidates finding, and later verify each candidate by spatial relations among these features, while the knowledge-based methods usually exploit information of the whole image and are sensitive to complicated backgrounds.

Face Detection Using Color Information

In this work, Hsu et al [7] proposed to combine several features for face detection. They used color information for skin-color detection to extract candidate face regions. In order to deal with different illumination conditions, they extracted the 5% brightest pixels and used their mean color for lighting compensation. After skin-color detection and skin-region segmentation, they proposed to detect invariant facial features for region verification. Human eyes and mouths are selected as the most significant features of faces and two detection schemes are designed based on chrominance contrast and morphological operations, which are called "eyes map" and "mouth map". Finally, we form the triangle between two eyes and a mouth and verify it based on (i) luminance variations and average gradient orientations of eye and mouth blobs, (ii) geometry and orientation of the triangle, and (iii) the presence of a face boundary around the triangle. The regions pass the verification are denoted as faces and the Hough transform are performed to extract the best-fitting ellipse to extract each face. This work gives a good example of how to combine several different techniques together in a cascade fashion. The lighting compensation process doesn't have a solid background, but it introduces the idea that despite modelling all kinds of illumination conditions based on complicated probability or classifier models, we can design an illumination-adaptive model which modifies its detection threshold based on the illumination and chrominance properties of the present image. The eyes map and the mouth map show great performance with fairly simple operations and in our recent work we also adopt their framework and try to design more robust maps.

Face detection based on random labelled graph matching

Leung et al. developed a probabilistic method to locate a face in a cluttered scene based on local feature detectors and random graph matching [8]. Their motivation is to formulate the face localization problem as a search problem in which the goal is to find the arrangement of certain features that is most likely to be a face pattern. In the initial step, a set of local feature detectors is applied to the image to identify candidate locations for facial features, such as eyes, nose, and nostrils, since the feature detectors are not perfectly reliable, the spatial arrangement of the features must also be used for localize the face. The facial feature detectors are built by the multi-orientation and multi-scale Gaussian derivative filters, where we select some characteristic facial features (two eyes, two nostrils, and nose/lip junction) and generate a prototype filter response for each of them. The same filter operation is applied to the input image and we compare the response with the prototype responses to detect possible facial features. To enhance the reliability of these detectors, the multivariate-Gaussian distribution is used to represent the distribution of the mutual distances among each facial feature, and this distribution is estimated by a set of training arrangements. The facial feature detectors averagely find 10-20 candidate locations for each facial feature, and the brute-force matching for each possible facial feature arrangement is computationally very demanding. To solve this problem, the authors proposed the idea of controlled search. They set a higher threshold for strong facial feature detection, and each pair of these strong features is selected to estimate the locations of other three facial features using a statistical model of mutual distances. Furthermore, the covariance of the estimates can be computed. Thus, the expected feature locations are estimated with high probability. Constellations are formed only from candidate facial features that lie inside the appropriate locations, and the ranking of constellation is based on a probability density function that a constellation corresponds to a face versus the probability it was generated by the non-face mechanism. In their experiments, this system is able to achieve a correct localization rate of 86% for cluttered images.

Template matching methods

In this category, several standard patterns of a face are stored to describe the face as a whole or the facial feature separately. The correlations between an input image and the stored pattern are computed for detection. These methods have been used for both face localization and detection. The following subsection summarizes an excellent face detection technique based on deformable template matching, where the template of faces is deformable according to some defined rules and constraints.

Adaptive appearance model

In the traditional deformable template matching techniques, the deformation constraints are determined based on user-defined rules such as first- or second-order derivative properties. These constraints are seeking for the smooth nature or some prior knowledge, while not all the patterns we are interested in have these properties. Furthermore, the traditional techniques are mainly used for shape or boundary matching, not for texture matching.

V. COMPARISON AND CONCLUSION

In this section, we'll give summaries on face detection and face recognition techniques during the past twenty year as well as popular face data set for experiments and their characteristics.

Method	Category	Characteristics
Hierarchical knowledge-based	Knowledge-based	Coarse-to-fine procedure
Horizontal / vertical projection	Knowledge-based	
Face Detection Using Color Information	Feature-based	Combining skin-color detection, face shape verification, and facial feature configuration for detection
Face detection based on random labelled graph matching graph matching [11]	Feature-based	Combining simple features with statistical learning and estimation
Active appearance model	Template matching	Learning facial shape and appearance variation by data
Example-based learning	Appearance-based	Learning the face and non-face distribution by mixture of Gaussian
Haar features with Ad boost	Appearance-based	Ad boost for speed-up
Generative models	Part-based	Unsupervised extracting important facial features, and learning the relation among parts and discrimination
Component-based with SVM	Part-based	Learning global and local SVM for detection

Table 2: The summary of face detection techniques

Method	Category	Characteristics
PCA	Holistic-based	PCA for learning Eigen faces, unsupervised
LDA	Holistic-based	LDA for learning fisher faces, supervised
2D-PCA	Holistic-based	2D-PCA for better statistical properties
ICA	Holistic-based	ICA for catch facial independent components, two architectures are proposed
Laplacian faces	Holistic-based	Nonlinear dimension reduction for finding bases, LPP
Evolutionary pursuit	Holistic-based	Using the genetic algorithm for finding the best projection bases based on generalization error
Kernel PCA And Kernel LDA	Holistic-based	Mapping the image into higher-dimensional space by the kernel function, and exploit the PCA and LDA bases there
Sparse representation	Holistic-based	Using L1 minimization and over-complete dictionary for finding sparse representation
Gabor and dynamic link architecture	Feature-based	Gabor features extracted at facial feature locations, while performing one-by-one matching
Gabor and elastic bunch graph matching	Feature-based	Gabor features extracted at facial feature locations, and obtaining the robust representation by the FBG matching.
LBP	Feature-based	Local binary patterns are introduced
LTP	Feature-based	Binary into ternary
AAM	Template matching	AAM parameters for classification learning
Component-based	Part-based	Comparing global and component representation,
SIFT	Part-based	Using SIFT feature with spatial constraints to

Table 3: The summary of face recognition techniques

Name	RGB/gray	Image size	# people	Pictures/person	Conditions	Available
AR Face Database*	RGB	576x768	126	26	i, e, o, t	Yes
Richard's MIT	RGB	480x640	154	6	p, o	Yes
CVL	RGB	640x480	114	7	p, e	Yes
The Yale Face Database B*	Gray	640x480	10	576	p, i	Yes
The Yale Face Database*	Gray	320x243	15	11	i, e	Yes
PIE*	RGB	640x486	68	~608	p ,i, e	Yes
The UMIST Face Database	Gray	220x220	20	19-36	p	Yes
Olivetti Att-ORL*	Gray	92x112	40	10		Yes
JAFFE	Gray	256x256	10	7	e	Yes
The Human Scan	Gray	384x286	23	~66		Yes
XM2VTSDB	RGB	576x720	295		p	With pay
FERET*	RGB/gray	256x384	30000		p, i, e, i/o, t	Yes

Table 4: The summary of popular databases used for detection and recognition tasks

The (*) points out most used databases. Image variations are indicated by (i) illumination, (p) pose, (e) expression, (o) occlusion, (i/o) indoor/outdoor conditions and (t) time delay.

VI. ACKNOWLEDGEMNT

It is great opportunity for me to write about the topic like "Facial Detection and Recognition: A Detailed Survey" so I like to acknowledge my gratitude to Mr. Sanjay Singh, Head of Computer Science and Engineering Department, Aravali College Of Engineering & Management who provided me this opportunity and also guide me to towards the correct path. Last but not the least, I would like to thanks to my family for their support and encouragement.

REFERENCES

[1] [online] Available: http://www.michaelbach.de/ot/fcs_thompson-thatcher/index.html.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7,711–720, 1997.

[3] J. Yang D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 1, 131–137, 2004.

[4] G. Yang and T. S. Huang, "Human face detection in complex background," Pattern Recognition Letter, vol. 27, no. 1, pp. 53-63, 1994.

[5] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting face in images: a survey," IEEE Trans. Pattern Analysis and Machine Intelligence vol.24, pp-34-58.

[6] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views," Proc. Int'l Conf. Acoustics, Speech and Signal Processing, vol. 4, pp. 2537-2540, 1997.

[7] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain," Face detection in color images," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, 696–706, 2002.

[8] T. K. Leung, M. C. Burl, and P. Perona, "Finding faces in cluttered scenes using random labelled graph matching," Proc. Fifth IEEE Int'l Conf. Computer Vision, pp. 637-644, 1995.

[9] P. Viola and M. Jones, "Robust real-time face detection," Int'l Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.[10] R. O. Duda, P. E. Hart, D. G. Stoke, Pattern classification, 2nd ed., John Wiley & Sons, 2001.

[11] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer,2005.

[12] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," IEEE Conf. Computer Vision and Pattern Recognition, pp. 657–662, 2001.

[13] M. H. Yang, ""Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods," AFGR, pp. 205–211, 2002.

[14] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 6, pp. 570–582, 2000.

[15] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer,2005.

[16] S. Theodoridis and K. Koutroumbas, Pattern recognition, 4th ed., Academic Press, 2009.

[17] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," IEEE Trans. Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.

[18] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacian faces," IEEE Trans Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328-340, 2005.

[19] X. He and P. Niyogi, "Locality Preserving Projections," Proc. Conf. Advances in Neural Information Processing Systems, 2003.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, pp. 2323-2326, 2000.

[21] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face recognition using active appearance models," Proc. European Conf. Computer Vision, vol. 2, pp. 581-695, 1998.

[22] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face Recognition: Component-based versus global approaches," Computer Vision and Image Understanding, vol. 91, nos. 1-2, pp. 6-21, 2003.

[23] R. Rifkin, Everything old is new again: a fresh look at historical approaches in machine learning, Ph.D. thesis, M.I.T., 2002.

[24] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. L. Lu, "Person-Specific SIFT Features for Face Recognition," Int'l Conference on Acoustic, Speech, and Signal Processing, 2007.

[25] D. G. Lowe, "Distinctive image features from scale-invariant key points," Int'l Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[26] Sanjay Singh, Sandeep Gupta, "Study of Evolutionary Algorithm Application: Information Retrieval", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN: 2349-5162, Vol.6, Issue 4, page no.244-252, April-2019, Available: http://www.jetir.org/papers/JETIR1904436.pdf

[27] Sanjay Singh, Sandeep Gupta, "Information Storage using Quantum entangled states: Image Optimization", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN: 2349-5162, Vol.6, Issue 4, page no.338-342, April-2019, Available:http://www.jetir.org/papers/JETIR1904551.pdf

[28]A.FischlerandR.A.Elschlager,"TheRepresentationandMatchingofPictorialstructures",Lockheed Palo Alto Research Lab · LockheedMissiles&SpaceCompany,Inc.Cal.,LMSC-D243781, Sept.1971.

[29] J.Goldstein,, L.D.Harmonand and A.B. Lesk:"Identification of Human Faces",Proc. IEEE,Vol.59,p.748,1971.