

Information Theory and Neural Network based Approach for Face Recognition: A Review

S.N. Kakarwal¹, Dr. R.R. Deshmukh²

¹ P.E.S. College of Engineering, Department of Computer Science and Engineering, Aurangabad, India
Email: s_kakarwal@yahoo.com

² Dr. Babasaheb Ambedkar Marathwada University, Department of Computer Science and Information Technology, Aurangabad, India
Email: ratnadeep_deshmukh@yahoo.co.in

Abstract— In face recognition, it is important to select the invariant facial features especially faces with various pose and expression changes. This paper presents novel feature extraction techniques such as Entropy and Mutual Information and for classification Feed forward neural network is used, which will be better than traditional methods for accurately recognizing the faces.

Index Terms— Biometrics, Information Theory, Entropy and mutual information and Feed forward neural network.

I. BIOMETRICS

Biometric can be defined as technique of studying physical characteristics of a person such as fingerprints, hand geometry, eye structure etc. to establish his or her identity. Biometrics-based personal identification techniques that use physiological or behavioral characteristics are becoming increasingly popular compared to traditional token-based or knowledge based techniques such as identification cards (ID), passwords, etc. One of the main reasons for this popularity is the ability of the biometrics technology to differentiate between an authorized person and an impostor who fraudulently acquires the access privilege of an authorized person [1]. Popular pattern recognition paradigms based on data reduction, such as redundancy reduction and dimensionality reduction, have met with difficulties in solving complex pattern recognition problems, such as the human face recognition problem [2].

A brief description of some commonly used biometrics is given below:

A. Face

Face recognition is a non-intrusive method, and facial images are probably the most common biometric characteristic used by humans to make personal recognition. The most popular approaches to face recognition are based on either: 1) the location and shape of facial attributes, such as the eyes, eyebrows, nose, lips, and chin and their spatial relationships or 2) the overall (global) analysis of the face image that represents a face as a weighted combination of a number of canonical faces. While the authentication performance of the face recognition systems that are commercially available is reasonable, they impose a number of restrictions on how the facial images are obtained, often requiring a fixed and simple background or special illumination. These systems also have difficulty in matching face images captured

from two drastically different views and under different illumination conditions (i.e., varying temporal contexts).

B. Fingerprint

Humans have used fingerprints for personal identification for many decades and the matching (i.e., identification) accuracy using fingerprints has been shown to be very high. A fingerprint is the pattern of ridges and valleys on the surface of a fingertip, the formation of which is determined during the first seven months of fetal development.

One problem with the current fingerprint recognition systems is that they require a large amount of computational resources, especially when operating in the identification mode. Finally, fingerprints of a small fraction of the population may be unsuitable for the automatic identification because of genetic factors, aging, environmental, or occupational reasons [3].

C. Retina

The retinal vasculature is rich in structure and is supposed to be a characteristic of each individual and each eye. It is claimed to be the most secure biometric since it is not easy to change or replicate the retinal vasculature. The image acquisition requires a person to peep into an eye-piece and focus on a specific spot in the visual field so that a predetermined part of the retinal vasculature could be imaged. The image acquisition involves cooperation of the subject, entails contact with the eyepiece, and requires a conscious effort on the part of the user. All these factors adversely affect the public acceptability of retinal biometric [4].

In the Table I comparison of various Biometric Techniques is given [5]:

TABLE I
COMPARISON OF VARIOUS BIOMETRIC TECHNIQUES

Rank	Accuracy	Convenience	Cost
1	DNA	Voice	Voice
2	Iris	Face	Signature
3	Retina	Signature	Finger
4	Finger	Finger	Face
5	Face	Iris	Iris
6	Signature	Retina	Retina
7	Voice	DNA	DNA

The rest of this paper is organized as follows. Section II explain two feature extraction techniques namely entropy and mutual information. In Section III, classification method is briefly described, which is devoted to Feed forward neural network algorithm. In section IV the application of face recognition system is given. Conclusion and discussions are outlined in the last section.

II FEATURE EXTRACTION

Broadly speaking, features are any extractable measurement used. Examples of low-level features are signal intensities. Features may be symbolic, numerical or both. An example of a symbolic feature is color; an example of numerical feature is weight.

The related problems of feature selection and feature extraction must be addressed at the outset of any pattern recognition system design. The key is to choose and to extract features that (1) are computationally feasible; (2) lead to ‘good’ PR system success and; (3) reduce the problem data into a manageable amount of information without discarding valuable information [6].

The feature extraction is used to reduce the dimension of the face space by transforming it into feature representation. It will be responsible for transforming or composing the normalized pixel values of the face image and represents it into an appropriate representation or feature vector by finding the key features that will be used for classification [7]. Typically, there are two primary goals for dimensionality reduction: (i) data compression and (ii) feature extraction for classification purposes. While PCA has been proven to be an optimal method for data compression, it is not necessarily an optimal method for feature extraction, particularly when the features are used in a supervised classifier [8].

This paper deals with two important feature extraction techniques entropy and mutual information.

A. Entropy and Mutual Information

Shannon gave a precise mathematical definition of the average amount of information conveyed per source symbol, which is termed as Entropy [9].

The Shannon mutual information has an advantage that it can be regarded as a measure of statistical independence of two random variables, that is, they are (statistically) independent iff the mutual information is equal to zero.

Consider two random variables and having some joint probability distribution over a finite set. The unconditional uncertainty of can be measured by different entropies, the most famous of which is the Shannon entropy. Some of them have been given practical interpretations, e.g., the Shannon entropy can be interpreted in terms of coding and the min entropy in terms of decision making and classification [10]

Entropy is a statistical measure that summarizes randomness.

Given a discrete random variable, its entropy is defined by

$$H(X) = -E_x[\log P(X)]$$

$$= - \sum_{x_i \in \Omega_x} P(X = x_i) \log(X = x_i) \tag{1}$$

Where Ω_x is the sample space and x_i is the member of it. $P(X=x_i)$ represents the probability when X takes on the value x_i . We can see in (1) that the more random a variable is, the more entropy it will have.

Conditional and joint entropy relate the predictability of two random variables. Conditional entropy is defined as

$$H(Y / X) = -E_x[E_y[\log P(Y / X)]]$$

$$= \sum_{x_i \in \Omega_x} \sum_{y_j \in \Omega_y} [P(X = x_i, Y = y_j) \times \log P(Y = y_j / x = x_j)] \tag{2}$$

Joint entropy is defined as

$$H(X, Y) = -E_x[E_y[\log P(X, Y)]]$$

$$= \sum_{x_i \in \Omega_x} \sum_{y_j \in \Omega_y} [P(X = x_i, Y = y_j) \times \log P(X = x_i, Y = y_j)] \tag{3}$$

Conditional entropy measures the randomness of Y when X is given, and joint entropy measures the randomness of Y and X. With the increase of $H(Y/X)$, Y gets more dependent on X. However, conditional entropy by itself is not a measure of dependency. A small value of $H(Y/X)$ implies either $H(Y)$ is small or Y is less dependent on X. When X and Y are independent, (2) and (3) can be expressed using $H(X)$ and $H(Y)$ [11].

$$H(Y / X) = H(Y) \tag{4}$$

$$H(X, Y) = H(X) + H(Y) \tag{5}$$

MI between two random variables is defined as

$$MI(Y, X) = H(Y) - H(Y / X) \tag{6}$$

Because conditional entropy can be expressed in terms of marginal and joint entropies

$$H(Y / X) = H(X, Y) - H(X) \tag{7}$$

We can get two equivalent expressions for MI

$$MI(X, Y) = H(X) - H(X / Y) \tag{8}$$

$$MI(X, Y) = H(X) + H(Y) - H(X / Y) \tag{9}$$

MI measures the statistical dependency between two random variables. The physical meaning of MI is the reduction in entropy of Y given X. This is best demonstrated in the (8). In (8), $H(X)$ is the entropy of X, computed on the probability distribution of X. $H(X/Y)$ denotes the conditional entropy, which is based on the conditional probability $P(X/Y)$. When interpreting entropy as a measure of uncertainty, (8) translates to “the amount of uncertainty about X minus the uncertainty X about when Y is known”. In other words, MI is the amount by which the uncertainty about X decreases when is Y given [11].

III. CLASSIFICATION TECHNIQUE

The task of the classifier component of a full system is to use the feature vector provided by the feature extractor to assign the object to a category. The degree of difficulty of the classification problem depends on the variability in the feature values for the objects in the same category relative to the difference between feature values for objects in the different categories [12].

Many classification techniques are available, in this paper we have focused feed forward neural network for the purpose of classification.

A. Feed Forward Net

In recent years, there has been an increase in the use of evolutionary approaches in the training of artificial neural networks (ANNs). While evolutionary techniques for neural networks have shown to provide superior performance over conventional training approaches, the simultaneous optimization of network performance and architecture will almost always result in a slow training process due to the added algorithmic complexity [13].

Feed forward networks may have a single layer of weights where the inputs are directly connected to the output, or multiple layers with intervening sets of hidden units. Neural networks use hidden units to create internal representations of the input patterns [14].

A Feed forward artificial neural network consists of layers of processing units, each layer feeding input to the next layer in a Feed forward manner through a set of connection weights or strengths. The weights are adjusted using the back propagation learning law. The patterns have to be applied for several training cycles to obtain the output error to an acceptable low value.

The back propagation learning involves propagation of the error backwards from the input training pattern, is determined by computing the outputs of units for each hidden layer in the forward pass of the input data. The error in the output is propagated backwards only to determine the weight updates [15]. FFNN is a multilayer Neural Network, which uses back propagation for learning [16].

IV. APPLICATION

Access control by face recognition has the following advantages in comparison with other biometrics systems. There are no requirements for expensive or specialized equipment; a system may be built using a simple video camera and a personal computer. The system is passive. There is no need to touch something by fingers or palm, no need to say any word or lean eye to a detector. Any person just may walk or stay before the camera, and the system performs recognition. It is especially useful in everyday usage. Also it has advantages in different extremal or non-standard situations, when it is impossible or inconvenient to take other biometric characteristics, for example when catching criminals [17].

Given an image, recognize the faces that are currently seen. The same face can be seen incredibly different in different images due to differences in view point, illumination etc. A robust recognizer must develop invariant recognition (translation, size and rotation). In theory, security systems involving face recognition would be impossible to hack, as the identification process involves unique identification methods, and thus only authorized users will be accepted. This mechanism would be convenient, with no need to remember passwords or personal identification numbers.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, we have proposed a novel feature extraction approach, entropy and mutual information to

extract features from each face. The derived features are later used to form a new feature vector, which will be used as the input to the classifier. The new algorithm will be able to perform much better than tradition methods like eigenface and fisherface.

Further research will focus on searching new entropy-based feature representation, and the design of neural network based classifiers.

ACKNOWLEDGEMENTS

The author would like to thank the anonymous reviewers for their critical and constructive comments and suggestions.

REFERENCES

- [1] Anil K. Jain and Umut Uludag, "Hiding Biometric Data", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 11, NOVEMBER 2003
- [2] Chengjun Liu, "Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 5, MAY 2006
- [3] Anil K. Jain, Arun Ross, and Sharath Pankanti, "Biometrics: A Tool for Information Security", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 1, NO. 2, JUNE 2006
- [4] Anil K. Jain, Arun Ross, and Salil Prabhakar, An Introduction to Biometric Recognition, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 14, NO. 1, JANUARY 2004, pp-1.
- [5] www.scribd.com/doc/3834277/Biometrics
- [6] Robert Schalkoff, Pattern Recognition, Wiley Publication, pp. no. 6-7.
- [7] Jamikaliza Abdul Karim Rubiyah Yusof Marzuki Khalid, "Investigate the performance of Fuzzy Artmap Classifier for Face Recognition System", 2008 IEEE International Conference on Signal Image Technology and Internet Based Systems 978-0-7695-3493-0/08 \$25.00 © 2008 IEEE
- [8] Anil Cheriyadat, Lori Mann Bruce, "Why Principal Component Analysis is not an Appropriate Feature Extraction Method for Hyperspectral Data", 0-7803-7929-2/03/\$17.00 (C) 2003 IEEE
- [9] Richard Wells, Applied Coding and Information Theory for engineers, Pearson Education, pp. no. 18
- [10] Jovan Dj. Golic and Madalina Baltatu, "Entropy Analysis and New Constructions of Biometric Key Generation Systems", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 54, NO. 5, MAY 2008
- [11] Lifeng Liu, Tianzi Jiang, Jianwei Yang, and Chaozhe Zhu, "Fingerprint Registration by Maximization of Mutual Information", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 5, MAY 2006
- [12] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, A Wiley-Interscience Publication, pp. no. 12
- [13] Chi-Keong Goh, Eu-Jin Teoh, and Kay Chen Tan, "Hybrid Multiobjective Evolutionary Design for Artificial Neural Networks", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 19, NO. 9, SEPTEMBER 2008
- [14] S.N. Sivanandanam, S. Sumathi, S. N. Deepa, Introduction to Neural Networks using MATLAB 6.0, Tat McGrawHill Publishing Company Limited, pp.20
- [15] B. Yegnanarayana, "Artificial Neural Networks", PHI
- [16] Sangeeta Kakarwal, Mahananda Malkauthekar, Shubhangi Sapkal, Dr. Ratnadeep Deshmukh, "Face Recognition Using FD and FFNN", 2009 IEEE International Advance Computing Conference (IACC 2009), pp 2740-2742
- [17] Dmitry Bryliuk and Valery Starovoitov, "Access Control By Face Recognition Using Neural Networks", УДК 681.3.01+681.327.12