A Novel and Effective Method on Studying the Success of Nearest Neighbor Methods in Prediction

Apoorva.A
Assistant professor, BCA department, New Horizon College Marathalli Bangalore.

Received: January 06, 2019 Accepted: February 12, 2019

ABSTRACT: *Many modern methods for prediction leverage nearest neighbor search to find past training examples most similar to a test example, an idea that dates back in text to at least the 11th century and has stood the test of time. This monograph aims to explain the success of these methods, both in theory, for which we cover foundational nonasymptotic statistical guarantees on nearest-neighbor-based regression and classification, and in practice, for which we gather prominent methods for approximate nearest neighbor search that have been essential to scaling prediction systems reliant on nearest neighbor analysis to handle massive datasets. Furthermore, we discuss connections to learning distances for use with nearest neighbor methods, including how random decision trees and ensemble methods learn nearest neighbor structure, as well as recent developments in crowd sourcing and graphons.*

***Key Words**:*

1. INTRODUCTION

1.1 Biometrics

Biometrics is used in the process of authentication of a person by verifying or identifying that a user requesting a network resource is who he, she, or it claims to be, and vice versa. It uses the property that a human trait associated with a person itself like structure of data with the incoming data we can verify the identity of a particular person [1]. There are many types of biometric system like detection and recognition, iris recognition etc., these traits are used for human identication in surveillance system, criminal identification , face details etc. By comparing the existing fingerprint recognition.

1.2 Face Recognition

Human beings have recognition capabilities that are unparalleled in the modern computing era. These are mainly due to the high degree of interconnectivity, adaptive nature, learning skills and generalization capabilities of the nervous system. The human brain has numerous highly interconnected biological neurons

which, on some specific tasks, can outperform super computers. The main idea is to engineer a system which can emulate what a child can do. Advancements in computing capability over the past few decades have enabled comparable recognition capabilities from such engineered systems quite successfully. Early face recognition algorithms used simple geometric models, but recently the recognition process has now matured into a science of sophisticated mathematical representations and matching processes. Major advancements and initiatives have propelled face recognition technology into the spotlight. Face recognition technology can be used in wide range of applications. Computers that detect and recognize faces could be applied to a wide variety of practical applications including criminal identification etc. Face recognition and detection can be achieved using technologies related to computer science. Features extracted from a face are processed and compared with similarly processed faces present in the database. If a face is recognized it is known or the system may show a similar face existing in database else it is unknown. In surveillance system if a unknown face appears more than one time then it is stored in database for further recognition. These steps are very useful in criminal identification. In general, face recognition techniques can be divided into two groups based on the face representation they use appearance-based, which uses holistic texture features and is applied to either whole-face or specific face image and feature-based, which uses geometric facial features and geometric relationships between them.(A few example applications are shown in Fig 1)

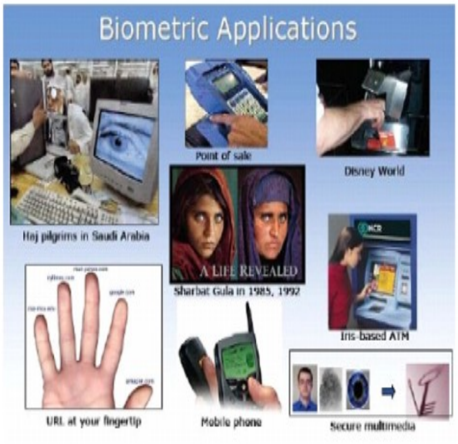


Figure 1: Biometric Applications

2. LITARATURE REVIEW

Several algorithms and techniques for face recognition have been developed in the past by researchers.These are discussed briefly in this section.

2.1 Face Recognition Based on Independent Component Analysis:

A number of current face recognition algorithms use face representations found by unsupervised statistical methods. Typically these methods find a set of basis images and represent faces as a linear combination of those images. Principal component analysis (PCA) is a popular example of such methods. The basis images found by PCA depend only on pairwise relationships between pixels in the image database. In a task such as face recognition, in which important information may be contained in the high-order relationships among pixels, it seems reasonable to expect that better basis images may be found by methods sensitive to these high-order statistics. Independent component analysis (ICA), a generalization of PCA, is one such method. We used a version of ICA derived from the principle of optimal information transfer through sigmoidal neurons. ICA was performed on face images in the FERET database under two different architectures, one which treated the images as random variables and the pixels as outcomes, and a second which treated the pixels as random variables and the images as outcomes. The first architecture found spatially local basis images for the faces. The second architecture produced a factorial face code. Both ICA representations were superior to representations based on PCA for recognizing faces across days and changes in expression. A classifier that combined the two ICA representations gave the best performance.[1]

2.2 Eigen-spaces:

Eigenspace-based face recognition corresponds to one of the most successful methodologies for the computational recognition of faces in digital images. Starting with the EigenfaceAlgorithm, different eigenspace-based approaches for the recognition of faces have been proposed. They differ mostly in the kind of projection method used (standard, differential, or kernel eigenspace), in the projection algorithm employed, in the use of simple or differential images before/after projection, and in the similarity matching criterion or classification method employed. The aim of this paper is to present an independent comparative study among some of the main eigenspace-based approaches. We believe that carrying out independent studies is relevant, since comparisons are normally performed using the implementations of the research groups that have proposed each method, which does not consider completely equal working conditions for the algorithms. Very often, a contest between the abilities of the research groups rather than a comparison between methods is performed. This study considers theoretical aspects as well as simulations performed using the Yale Face Database, a database with few classes and several images per class, and FERET, a database with many classes and few images per class.[2].

3. THEORY METHODOLOGY AND ALGORITHM

The previous sections illustrate different techniques and methods of face detection and recognition. Each category of method performs well in certain criteria and also has drawbacks as well. Systems with robustness and certain level of accuracy are still far away. Keeping in view case study the following architecture is proposed for the detection and recognition system. As discussed earlier that the robust system catering the needs of real world situation is a challenging task. The images will be scanned by scanner and stored into database. Again the image will be scanned and stored into the database. Now two images of the same candidate will be stored into the database. The first step is to select desired images from the database then for comparisons them the next step is to detect faces from each image. The next step is to recognize that images as of the same candidate or not.

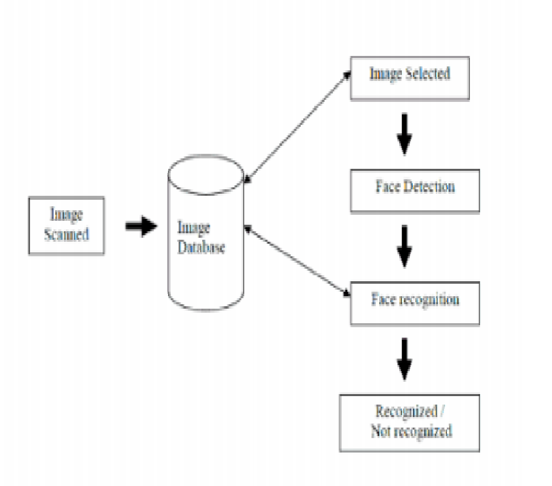


Figure 2: Structure of Face Recognition System

4. PROPOSED SYSTEM

The proposed face recognition system overcomes certain limitations of the existing face recognition system. It is based on extracting the dominating features of a set of human faces stored in the database and performing mathematical operations on the values corresponding to them. Hence when a new image is fed into the system for recognition the main features are extracted and computed to find the distance between the input image and the stored images. Thus, some variations in the new face image to be recognized can be tolerated. When the new image of a person differs from the images of that person stored in the database, the system will be able to recognize the new face and identify who the person is.

The proposed system is better mainly due to the use of facial features rather than the entire face. Its advantages are in terms of:

- Computational cost because smaller images (main features) require less processing to train the PCA.
- Recognition accuracy and better discriminatory power because of the use of color space conversion and dominant features and hence can be used as an effective means of authentication.
- Using KNN(nearest neighbour classifier)

The flowchart of the proposed method is shown in Fig2.

This method consist of three parts a) color-component feature selection with boosting, b) color FR solution using selected color component features, c)classifier using nearest neighbour classifier.

In this method the given image is transformed into various color space models like RGB, YCbCr, YIQ, HSV.Then the various color components available in the color space models are separated . For example from YCbCr image the Y,Cb,Cr components are splited separately. Then the eigen value and vector features are extracted from the various color component images using using PCA (Principal Component Analysis). Finally the extracted features are used to claissify test image by using Nearest Neighbour (NN) classifier.

4.1 METHODS

A. Color space models

A color space is a mathematical representation of a set of color. It deals with two fundamental color models, RGB (Red, Green, and Blue) used in color computer graphics and color television and YUV used in broadcast and television. Color spaces can be converted between each other, but video quality is lost with each conversion. Care should be taken to minimize the number of color space conversions used in the video encoding and decoding path, so as to loss as little as possible quality.

The red, green, and blue (RGB) color space is widely used throughout computer graphics and imaging. The RGB color space is the most prevalent choice for graphic frame buffers because color CRT's use red, green, and blue phosphors to create the desired color. All three RGB components need to be of equal bandwidth to generate any color within the RGB color cube.

The next color space model used in this paper is YCbCr. Here Y is luminance meaning that light intensity is non-linearly encoded using gamma correction. C_B and C_R are the blue-difference and red-difference chroma components.

The next color space model used in the paper is YIQ. In this model *I* stand for in-phase, while *Q* stands for quadrature, referring to the components used in quadrature amplitude modulation. The Y component represents the luma information; I and Q represent the chrominance information.

The next color space model used in the paper is HSV. HSV stands for hue, saturation, and value, and is also often called HSB (*B* for brightness) in each cylinder, the angle around the central vertical axis corresponds to "hue", the distance from the axis corresponds to "saturation", and the distance along the axis corresponds to "lightness", "value" or "brightness".

B. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical procedure that uses to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. Principal component analysis is appropriate when you have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables (called principal components) that will account for most of the variance in the observed variables.

Principal components analysis is a quantitatively rigorous method for achieving the simplification of data set. It is because, in data sets with many variables, groups of variables often move together. One reason for this is that more than one variable may be measuring the same driving principle governing the behaviour of the system. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other so there is no redundant information . The principal components as a whole form an orthogonal basis for the space of the data.

C. K-Nearest Neighbor Classifier

In pattern recognition, the K-nearest neighbor algorithm(K-NN) is a widely used classifier for classifying objects based on closest training examples in the feature space. The *k*-nearest neighbour algorithm is the simplest classifier of all machine learning algorithms. In this classifier image is classified by a majority vote of its neighbours. In kNN classifier the Euclidean distance between the testing image feature and each training image feature is determined to form a distance matrix. The summation value of diatnce matrix is estimated and sorted in increasing order .The first K elements are selected and majority class value is determined for classifying the image accurately.

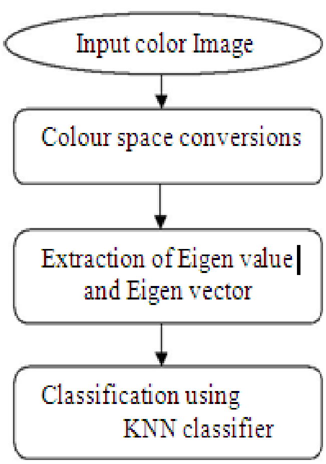


Fig.2 Flow chart of the proposed color Face Recognition framework

5.1 Advantages

1) High accuracy
2) Feature extraction for color image recognition
3) Supports dynamic face recognition

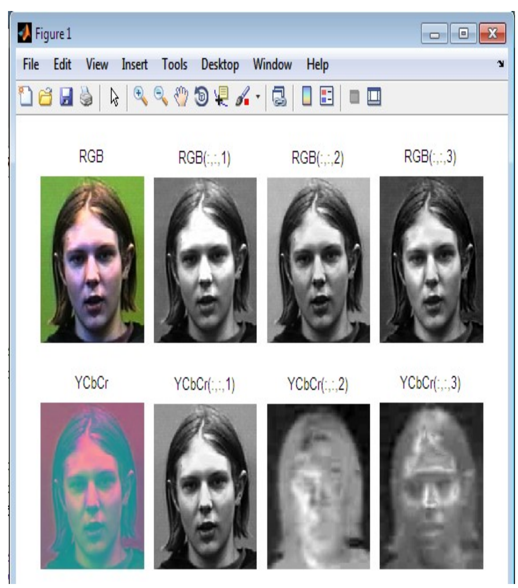
5.2 Applications

1. In Security
2. K-NN classifier is also better suitable for medical image classification
3. Office/Industries for authentication

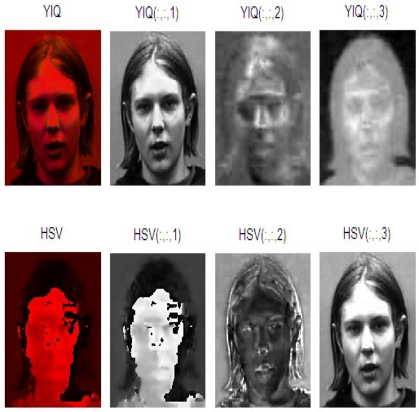
RESULTS AND DISCUSSIONS

The proposed method different persons face images are used to test the performance of the system. For Each person the database contains different facial expression images. face image features are considered as training feature. All the input images are converted into various color space models. The input image and various color components of the face image are shown in fig5.1 .The Eigen values and eigen vectors are extracted from each image. The summation value of eigen vector and eigen values are considered as features of each color component of an image.

(a) Input color image



(b)Color space components of RGB and YCbCr



(c)Color space components of YIQ and YIQ, and HSV color spaces

The KNN classifier is used to classify the different face images. In this project 70 label values are created for 70 persons. The Euclidean distance between the testing image feature and the training image feature is determined by finding the difference between the testing and the training feature and a distance matrix is created. In the distance matrix first k values are considered and the majority label of the k value is considered as the correct label of the given testing image. The performance of the system is measured in terms of accuracy. The accuracy is given by

$$Accuracy = \frac{correctly\ detected\ face\ Images}{total\ number\ of\ face\ images}$$

In the classification problem 700 image feature values are considered as training features and 350 image features are used for testing the classifier. Out of 700 training images 654 images are correctly classified and out of 350 testing image 317 images are are correctly classified without error .So the overall accuracy obtained by this method is 92.47%

6. CONCLUSION AND FUTUREWORK

In this project, a novel and effective color FR method is proposed. The image is transformed into various color space models. Then applying the PCA technique to every color space component and the features are extracted. The features are used by KNN classifier for recognizing the input face image. The accuracy obtained in this method is much better than other results available in the literature. In future the optimal set of color features is found out from the color space for increasing the accuracy.

7. REFFERENCES:

[1] P. Shih and C. Liu, "Improving the face recognition grand challenge baseline performance using color configurations across color spaces," in Proc. IEEE Int. Conf. Image Process., 2006, pp. 1001–1004.

[2] J. Wang and C. Liu, "Color image discriminant models and algorithms for face recognition," IEEE Trans. Neural Netw., vol. 19, no. 12, pp.2088–2097, Dec. 2008.

[3] C. Liu, "Learning the uncorrelated, independent, and discriminating color spaces for face recognition," IEEE Trans. Inf. Forensics ecurity, vol. 3, no. 2, pp. 213–222, Jun. 2008.

[4] Z. Liu and C. Liu, "A hybrid color and frequency features method for face recognition," IEEE Trans. Image Process., vol. 17, no. 10, pp. 1975–1980, Oct. 2008.

[5] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color face recognition for degraded face images," IEEE Trans. Syst.

[6] Man, Cybern. B, Cybern., vol. 39, no. 5, pp. 1217–1230, Oct. 2009.

[7] M. Rajapakse, J. Tan, and J. Rajapakse, "Color channel encoding with NMF for face recognition," in Proc. IEEE Int. Conf. Image Process.,2004,pp.2007–2010.

[8] J. Yang, C. Liu, and L. Zhang, "Color space normalization: Enhancingthe discriminating power of color spaces for face recognition," Pattern Recognit., vol. 35, no. 1, pp. 615–625, 2010.

[9] R. Likac and K. N. Plataniotis, Color Image Processing: Methods and Application. Boca Raton, FL: CRC Press, 2007.

[10] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 371–381, Mar. 2007.

[11] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 696–796, May 2002.

[12] A. Mura, "Upper bounds for error rates of linear combinations of classifiers," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 591–602, May 2002.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, 1997.

[14] R. E. Schapire, "The boosting approach to machine learning: An overview," in Proc. MSRI Workshop Nonlinear Estimation and Classification, 2002, pp. 149–172.

[15] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," IEEE Trans. Image Process., vol. 18, no. 8, pp. 1885–1886, Aug. 2009.

[17] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Using color local binary pattern features for face recognition," in Proc. IEEE Int. Conf. Image Process., 2010, pp. 4541–4544.