www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 7 Issue 5 May 2018, Page No. 23945-23951
Index Copernicus Value (2015): 58.10, 76.25 (2016) DOI: 10.18535/ijecs/v7i5.11

Estimation of age groups using facial recognition features

Suman Mann, Bhuvnesh Mehra , Anshul Mangla, Shubham Runwal, Prateek Singhal

[1,2,3,4,4]Dept of Information Technology, Maharaja Surajmal Institute of Technology, New Delhi-110058

Abstract-
Facial recognition has been used to detect emotions, gender, expressions and identity. These topics have been extensively studied. But, automatic Age Estimation is a topic that has not been researched much. The basis of our topic is that there are features on the human face that change as our age increases and in our project we utilize these features for age group classification. We classify the images into 6 age groups- (0-6, 8-12, 15-20, 25-32, 38-53, 60-100). The process involves 3 stages:-Pre-processing, Feature Extraction, Classification. Preprocessing includes commotion decreasing, standardization and change of the crude information into structure that is suitable for the pattern recognition. A small set of good features is selected from the available features, in order to provide best matching information. This process discovers important features to obtain an effective and improved answer for a given problem which is the next step. At last the classification or clustering stage is performed. Classification comprises of allotting a class mark to a given pattern while clustering discoveries homogeneous small groups in information. We also deal with comparison study of various feature extraction techniques (Haar feature extraction, HOG feature extraction) and classification techniques (Naïve Bayes, SVM, KNN, Neural Network-Back propagation algorithm) and finally try to determine the best possible combination for age group classification. Convolutional Neural Networks is also applied to estimate the age groups.

Keywords - face detection, age estimation, wrinkle features, pattern recognition.

Introduction

There is a vitale role of the face recognition in the law enforcement and security, despite of growing interest and advancement in this field, it is still a challenging problem, because determing the age of a person depends on various factors like expression, lifestyle etc.

In this research a successful age classification system is proposed with the help of face elements like surface and shape from human face pictures. The facial features can be extracted using Viola Jones Algorithm (Haar features) or HOG features and the classification of age can be done via Naïve Bayes, SVM, KNN or Neural Networks, the approach using CNN is also studied. This explores all the nine algorithms and their combinations and provide the best CNN approach which provides maximum accuracy.

Image Recognition is generally done in three steps. In the first step, information is collected by the use of few sensors or any other means and after that these unpolished information may be preprocessed. Preprocessing may include commotion decreasing, standardization and change of crude information into suitable structure for pattern recognition. In order to provide best matching information, a small set of good features is selected from the available features. This process discoveries important feature to obtain an effective and improved answer for a given problem which is the next step. Classification comprises of allotting a class mark to a given pattern while clustering discoveries homogeneous small groups in information

Face Recognition Methodology

Age group estimation is divided into two parts: first part which provides a comparison study of feature extraction (Haar feature extraction, HOG) and classification techniques (KNN, Neural Network algorithm) and then in second part the new and accurate methodology for age group estimation(that is combination of HOG and Neural Network) which is derived from accuracy of the comparison study.

Entire process is divided into three stages:

1. Image Preprocessing

2. Feature Extraction
3.Classification

The flow Graph below depicts the basic flow of the entire system.

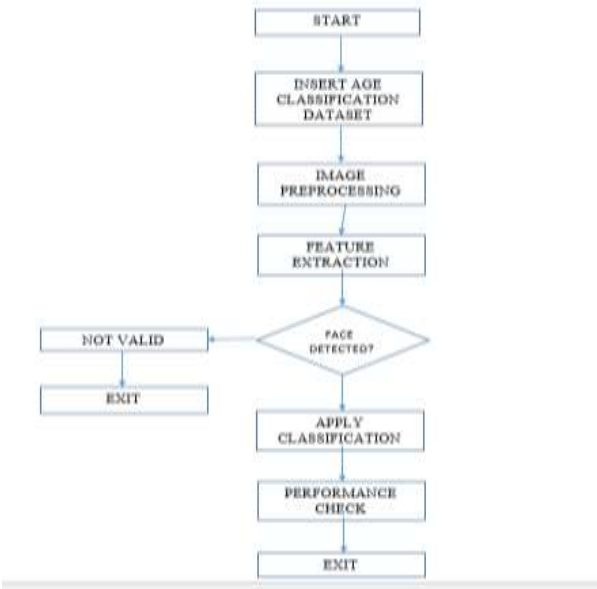


Fig1. Face Recognition Methodology

1. **Image Preprocessing**

Image Preprocessing is done to bring the image data to the lowest level of data abstraction. It's aim is to suppress the distortion in image and enhance some image features that are vital for further processing.

This phase involves applying various non-linear morphological operations to the given image because there is noise in the background while collecting the image. First the image is binarzied from RGB scheme and then converted into grayscale. After gray scaling, based on a experimental (dynamic) threshold, the photos are converted into black and white so that there is a uniformity in the images and no inaccuracy occurs due to skin color differences.

Feature Extraction

Feature Extraction is mainly done via two methods

1. Haar Features
2. OG (oriented gradient)

2.a Haar Features

Haar features have been used in object recognition, specifically for human face detection. In general, these feature add up the intensity values of the pixels in each rectangular region all inside an image subsection and computes the difference between the intensity values sum. The difference is then used to identify subsections of the image. For example, among all human faces the region of eyes is darker than the region of cheeks. Hence, a Haar-like feature for human face can be gathered by placing two adjacent rectangles over the eye and the cheek region. [14]

The major advantage of Haar features is its fast speed calculation, which makes it quite competitive in real-time object recognition.

Using Viola Jones Algorithm Haar like features are calculated in which the features are dived in to two categories i.e. Wrinkle features and Geometric features.

Geometrical Features

As desıgnated by the research of facıal representatıon and emotıonal cosmos, there ıs lot of change in the face features of humans as the age increases. In calculatıng geometrıcal features, global features in amalgamatıon wıth the grıd features are pull out from the face images. Followıng are the global features: the dıstance between two eye balls, dıstance between chın to eye, dıstance between nose tıp to eye and dıstance between eye to lıp.

By making use of four distance values as stated above, four features (namely Fea1, Fea2, Fea3 and Fea4) are calculated as mentioned below:

Fea1 = (distance between left eye ball and right eye ball) / (distance between eye and nose).

Fea2 = (distance between left eye ball and right eye ball) / (distance between eye and lip).

Fea3 = (distance between eye ball and nose) / (distance between eye and chin).

Fea4 = (distance between eye ball and nose) / (distance between eye and lip).

New borns have a large number of wrınkles on theır face. Addıtıonally the ratıo of new features is different from those in other lıfe spans. Hence we can say that it is more appropriate to use geometrıc features as compared to wrınkle features in order to determıne whether the image is of a baby or not [8].

In the case of a new born, the head is close to a cırcle. The dıstance between the two eyes balls is nearly equal to the distance between the eyes ball and the mouth. As the baby grows, the head bone also grows, the head grow into oval shaped and consequently there occurs a hasty rise in the

dıstance between eye ball to the mouth. Fınally, ratıo amid distance between baby's eye ball and noses is equal to the distance amid noses and mouths which is in turn are almost equal to one whereas as in case of the adults this distance is larger than 1.

Wrinkle Feature

The wrinkle feature has most ımportant property is the density of pıxels which helps to determıne the estimate age of any person. Estimation of this feature (Fea5) can be done in the following way:

Fea5= (summation of pixels present in forehead region / number of pixels present in forehead region) + (summation of pixels present in left eyelid region / number of pixels present in left eyelid region) + (summation of pixels in right eyelid region / number of pixels in right eyelid region) + (summation of pixels present in corner of left eye / number of pixels present in corner of left eye) + (summation of pixels present in corner of right eye / number of pixels present in corner of right eye).

Fea5 can be easily assessed with the help of grid features present in the face image that needs wrinkle geography present in image of particular face [10].

For the estimation of wrinkle features, steps are as follows :

As the age of a person increases, wrinkles on face become more visible and clearer. Aged individuals have vibrant wrinkles on the below mentioned areas of face [5]:

a) The forehead develop horizontal lines.
b) The corners of eyes develop wrinkles.
c) The cheeks have visible cheekbones, and clear lines between the cheeks and the upper lips.

Since there are apparent changes in wrinkle intensities and also some form of clear lines, hence Sobel edge magnitudes are used which help in approximating gradient magnitudes hence determining the level of wrinkles.

2.b Oriented Gradients (OG)

The algorithm can be described in this way that firstly, the images are divided into finest possible regions which are called as cells. For each selected cell, a histogram of gradient directions or edge directions is computed.Next each of the cell is parted into equivalent angular bins in agreement with its gradient direction.The calculated weighted gradient of cell contributes to its particular angular bin. The neighboring cells with same gradient direction are assembled together and the regions generated are known as block These alignments into blocks are the origin for histograms' normalization. The regularized group characterizes the block of histogram which is finally represented as the descriptor.

The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing.

The OG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. The OG descriptor is thus particularly suited for human detection in images.

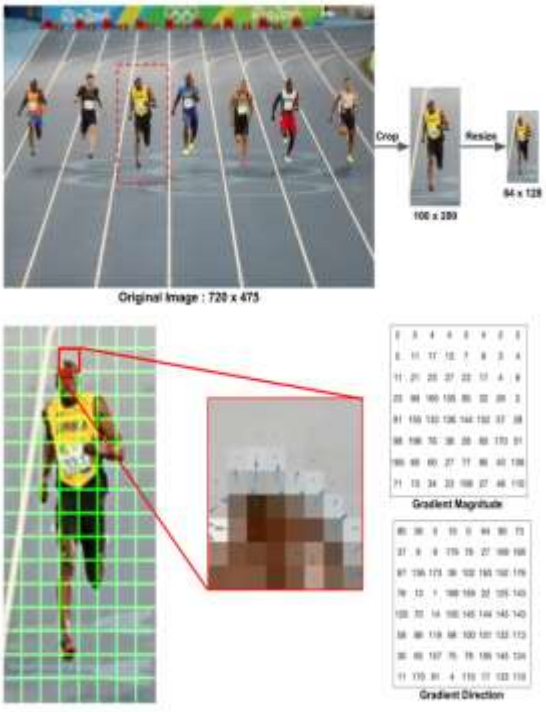


Fig 2. Oriented Gradients

3. Classification

The classification has been done using K nearest neighbours and neural networks. K nearest neighbours is a simple algorithm that stores all

available cases and classifies new cases based on a similarity measure (e.g., distance functions). While neural network is a simulated representation that is designed on the basis of human brain.

3.a Naïve Bayes

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods

A Naïve Bayes classifier is based on Bayes rule and simple form of Bayesian network. It is a simple probabilistic model relies on the assumption of feature independent in order to classify data. The algorithm assumes that each feature is independent of presence or absence of other feature in input data as this assumption is known as 'naïve'.

Bayes classifier use Bayes theorem, which is

$$P(c|F) = \frac{P\left(\frac{F}{c}\right)P(c)}{P(F)}$$

P(c/F) = probability of feature F being in classc

P(F/c) = probability of generating feature F given by class c

P(c) = probability of occurrence of class c

P(F) = probability of feature F occurring

In above context, we are looking for class c, so we find probable class given for given feature, F. Denominator does not depend on class so we treat it as a constant. Numerator depends on class so we focus to determine the value of P(F/c). For a class cj, the features are conditionally independent of each other, hence

$$P\left(f1, f2 \ldots \frac{fn}{cj}\right) = \prod_i P\left(\frac{fi}{cj}\right)$$

The Naïve Bayes is very simple and its conditional independence assumptions are not realistic in real world.

3.b Simple Vector Machines

The *one vs one* strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives. This strategy requires the base classifiers to produce a real-valued confidence score for its decision, rather than just a class label; discrete class labels alone can lead to ambiguities, where multiple classes are predicted for a single sample.

In pseudocode, the training algorithm for an OvO learner constructed from a binary classification learner *L* is as follows:

Inputs:

- *L*, a learner (training algorithm for binary classifiers)
- samples *X*
- labels *y* where $y_i \in \{1, \ldots K\}$ is the label for the sample X_i

Output:

- a list of classifiers f_k for $k \in \{1, \ldots, K\}$

Procedure:

- For each *k* in {1, …, *K*}
- Construct a new label vector *z* where z_i = 1 if $y_i = k$ and $z_i = 0$ otherwise
- Apply *L* to *X*, *z* to obtain f_k

Making decisions means applying all classifiers to an unseen sample *x* and predicting the label *k* for which the corresponding classifier reports the highest confidence score:

$$\hat{y} = \underset{k \in \{1 \ldots K\}}{\operatorname{argmax}} f_k(x)$$

This trains $K(K-1)/2$ binary classifiers for a *K*-way multiclass problem

I. IMPLEMENTATION

Age classification using feature selection (Haar and HOG) used two best classification techniques (neural network and KNN) in order to find the best solution for age classification.

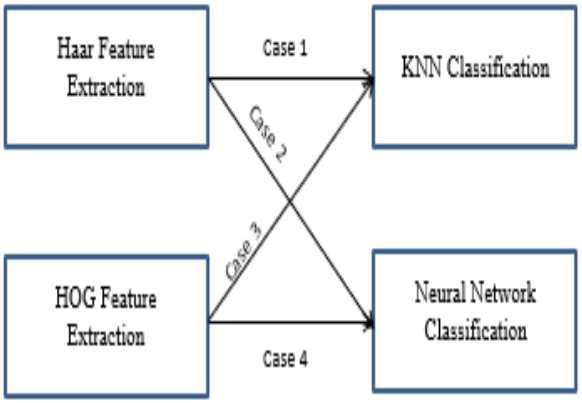


Fig 6. Four Cases

Case 1: Haar Feature Extraction and KNN Classification

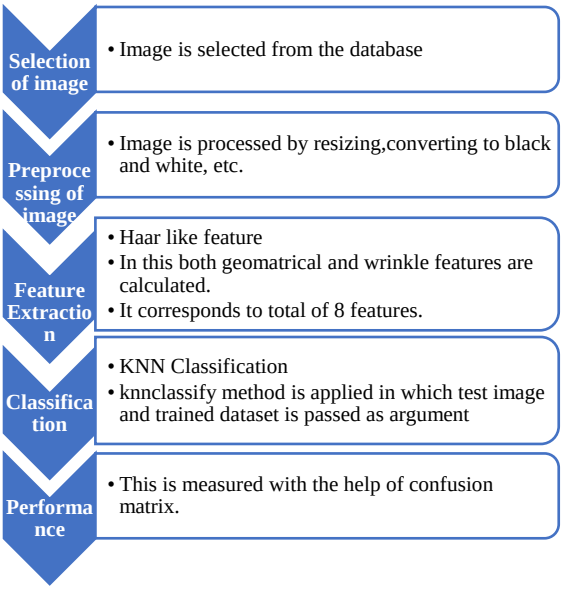
Case 2: Haar Feature Extraction and Neural Network Classification

Case 3: HOG Feature Extraction and KNN Classification

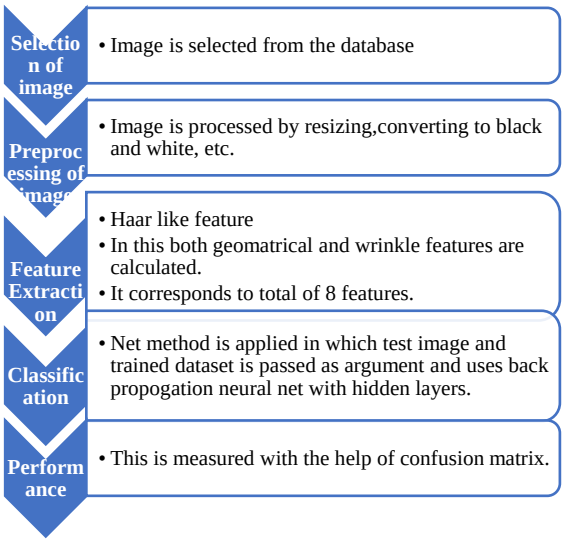
Case 4 HOG Feature Extraction and Neural Network Classification

These cases are explained with the help of flowcharts.

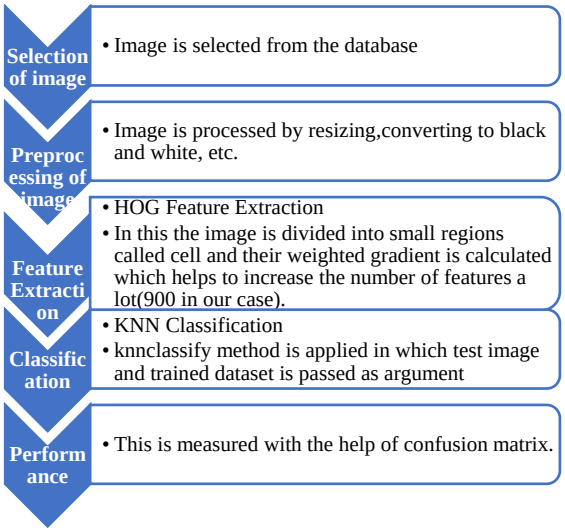
CASE 1: Haar Feature Extraction and KNN Classification



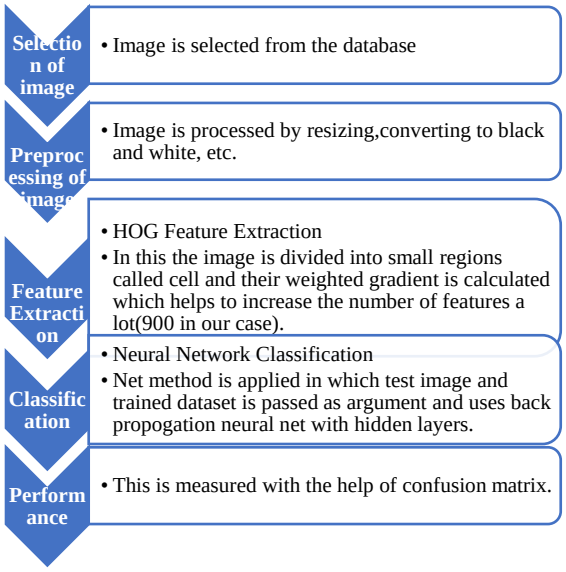
CASE 2: Haar Feature Extraction and Neural Network Classification



CASE 3: HOG Feature Extraction and KNN Classification



CASE 4: HOG Feature Extraction and Neural Network Classification



Results that we get from implementing these different combinations on the same data set is as follows:

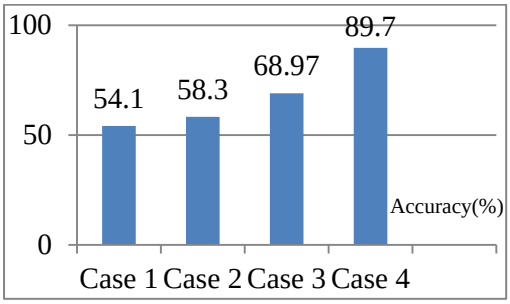


Fig 7. Accuracy Comparison

Conclusion

In this paper, a novel method for age group classification along with the comparative study of various techniques has been explained. Proposed technique first preprocesses the image which includes resizing the image, changing the image to black and white form from RGB format. Once the image is processed the features are extracted using Viola Jones algorithm (haar features), where both geometrical and wrinkle features are calculated providing a total of 8 features and HOG algorithm where gradients are calculated for each region of the face providing 900 features. Once the features are extracted two different classification algorithms (KNN and ANN) are applied which helps in classification of age groups. The observed results are noted and finally it is concluded that combination of HOG and ANN is the best methodology for age group estimation providing an accuracy of around 92%.

The features explored in HOG are also low which is done to decrease the timing constraint, once those features reach up to 4000, the accuracy of system will increase further more

References

[1] Abed, M. A., Ismail, A. N., & Hazi, Z. M. (2010). Pattern recognition using genetic algorithm. *International Journal of Computer and Electrical Engineering*,*2*(3), 1793-8163.

[2] A.K.Jain, "Age Invariant Face Recognition", IEEE Trans. On Pattern Analysis and Machine Intelligence, 2010

[3] A.Lanitis, C.Draganova, and C.Christodoulou, " Comparing different classifiers for automatic age estimation", IEEE Trans.Syst.Man, Cybern.B, Cybern, vol34, no.1m pp.621-628, Feb.2004.

[4] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*,vol. 24, no. 4, pp. 442–455, Apr. 2002.

[5] B.D., Zarit. B.J., Super, AND F.K.H. Quck, "Comparison of five color models in skin pixel classification", Int. Workshop on Recogntion , Analysis, and Tracking of faces and Gestures in Real Time Systems, pages 58-63, corfu, Greece, Sep. 1999.

[6] G. Guo, A. Jain, W. Ma, and H. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," IEEE Trans. Neural Netw., vol. 13, no. 4, pp. 811–820, Apr. 2002.

[7] G. Guo, S. Li, and K. Chan, "Face recognition by support vector machines," in Proc. IEEE Conf. FG, 2000, pp. 196–201..

[8] Golberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addion wesley, 1989.

[9] Jana, Ranjan, Debaleena Datta, and Rituparna Saha. "Age Group Estimation using Face Features." International Journal of Engineering and Innovative Technology (IJEIT) 3.2 (2013):130-134.

[10] Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press.

[11] Mahdy, Y. B., & El-Melegy, M. T. (1996, October). Encoding patterns for efficient classification by nearest neighbor classifiers and neural networks with application to handwritten Hindi numeral recognition. In Signal Processing, 1996.,

3rd International Conference on (Vol. 2, pp. 1362-1365). IEEE.
[12] Ma, H., Chan, J. C., Saha, T. K., & Ekanayake, C. (2013). Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources. Dielectrics and Electrical Insulation, IEEE Transactions on,20(2), 468-478.
[13] Tizhoosh, H. R. (2005). Image thresholding using type II fuzzy sets. Pattern recognition, 38(12), 2363-2372.
[14] X. F. He and P. Niyogi, "Locality preserving projections," presented at the NIPS, 2003.
[15] X.Geng, Z.H. Zhou, and K. Smith-Miles, " Automatic age estimation basend on facial aging patterns", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.29, pp.2234-2240, 2007.
[16] Y. Kwon and N. Lobo, "Age classification from facial images," Comput. Vis. Image Understand., vol. 74, no. 1, pp. 1–21, 1999.