Prediction of Human Ethnicity from Facial Images Using Neural Networks

Chapter *in* Advances in Intelligent Systems and Computing · June 2018
DOI: 10.1007/978-981-10-3223-3_20

5 authors, including:

Sarfaraz Masood
Jamia Millia Islamia

Shubham Gupta
AIIMS Bhopal All India Institute of Medical Sciences

Suhani Gupta
Jamia Millia Islamia

Musheer Ahmad
Jamia Millia Islamia

Prediction of human ethnicity from facial images using neural networks

Sarfaraz Masood[1], Shubham Gupta[2],Abdul Wajid[3], Suhani Gupta[4],

[1]Department of Computer Engineering, Jamia Millia Islamia,
New Delhi – 110025, INDIA, smasood@jmi.ac.in
[2,3,4]Department of Computer Engineering, Jamia Millia Islamia,
New Delhi – 110025, INDIA,
[2]shubham_0906@yahoo.com, [3]abdulw976@gmail.com, [4]suhanigupta139@gmail.com

Abstract. This work attempts to provide a solution for the problem of ethnicity classification of humans based on their facial features. This work is done for three major ethnicities: Mongolian, Caucasian and Negro .For this purpose, we took 447 sample images of FERET database, out of which 357 images were used for training and for 90 testing. Several geometric features and color attributes were extracted from the image and used for classification problem. The accuracy of the model obtained through artificial neural network is 82.4% whereas the accuracy obtained by deploying convolution neural network is 98.6%.

Keywords: ethnicity identification, artificial neural networks, convolutional neural networks, FERET Database.

1 Introduction

Face analysis is one of the most studied research topic in the field of computer vision and pattern recognition for the past few decades. Although face of a person provides a variety of demographic information like gender, age, ethnicity, etc yet ethnicity remains one of the invariant and fundamental attribute that cannot be easily masked like age and gender even in disguise. Therefore grouping people based on age and gender would not only complex the problem but will also yield wrong results. For this reason, ethnicity classification is a key component that can be deployed in various video surveillance systems at security checkpoints. Furthermore, this classification statement has potential application in image search query where prior knowledge of race would narrow down the search space in the database, thus simplifying the process.

In this work, three ethnic categories have been considered: The Mongolian, the Caucasian and the Negroid. As the people belonging to same category will have similar features. Similarly, people belonging to different ethnic category will have distinguishing features. This idea has helped us to extract and study fundamental features of each of these categories and classify them according to their distinguishing values.

(a) (b) (c)

Fig.1. Images of samples from (a)Negroid, (b)Caucasia and (c)Mongolian ethnicity

This work is organized into five sections: section II highlights the previous related research work and their contribution, section III describes the design of our algorithm and the methodology adopted, section IV deals with the results and discussions and section V finally concludes the paper.

2 Literature Review

There are a number of algorithms that have been devised over the years for the ethnicity identification of humans.

Work done by P. Viola and M. Jones [5] has provided the efficient and rapid method of detecting face in input image. This is a novel approach which uses Adaboost classifier. This has high detection rate with very less computation time on the dataset consisting of images under varying condition like illumination, pose, color, camera variation; etc. This algorithm is used in our approach to detect face in the image which will be later processed further.

Lu et al. [1] has proposed ethnicity classification algorithm in which image of the faces were examined at multiple scales .The Linear Discriminant Analysis (LDA) scheme is used for input face images to improve the classification result. The accuracy of the performance of this approach is 96.3% on the database of 2,630 sample images of 263 subjects. However, the dataset considered in this work consisted only of two classes i.e. Asian and non-Asian.

Hosoi et al. [2] have integrated the Gabor wavelet features and retina sampling for their work. These features were then used with the Support Vector Machines (SVM) classifier. This approach has used three categories: Asian, African and European. And the accuracy achieved for each category is: 96%, 94% and 93% respectively. However their approach seemed to have issues when considering other ethnicities.

In [3], ethnicity classification under the varying age and gender was performed on the very large scale dataset for the first time .The MORPHII dataset was used for this work which had 55,000 images. Guo and Mu has used Gabor features for classification problem of five ethnicities: Black, White, Hispanic, Asian and Indian. The prediction results for Black and White were good: 98.3% and 97.1% respectively. But due to insufficient dataset for other three races, prediction results deteriorated to 74.2% for Hispanic, 59.5% for Asian and 6.9% for Indian.

S. Md. Mansoor [4] has used Viola Jones [5] algorithm for face detection problem. After the detection of face, various features namely skin color; lip color and normalized forehead area were extracted from the image. This classification problem has used the Yale, FERET [6] dataset of Mongolian, Caucasian and Negroid images. The overall accuracy achieved in this work with these features was 81%.

It was evident from the above mentioned works that none had considered geometric features for their solution. Also the scope of pre-trained Convolution Neural Network has yet not been explored for this problem so far. Hence in this problem, we have considered geometric features for training the ANN, and have also attempted to use convolutional neural networks for solving this problem.

3. Design of Experiment

In an attempt to provide efficient solution to the ethnicity classification problem, we conducted two experiments. First experiment was done using Artificial Neural Network and the second one with Convolution Neural Network. The steps followed while carrying out these experiments are discussed below.

3.1 Experiment#1 (Artificial Neural Network)

In this experiment, various facial features were extracted from the sample images followed by training of a feedforward artificial neural network using a backpropagation based training algorithm. The face area is detected from the target image using Viola Jones [5] algorithm.

3.1.1 Geometric Feature Calculation

Once the face has been detected, using the cascade classifiers other facial areas like Nose, mouth, left and right eye were marked. Then the distances and ratios between these facial areas were calculated. The combinations of different geometric features of face were different for different ethnicities.

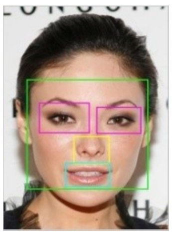

Fig. 2. Face and facial feature detection using Viola Jones.

Table 1 shows the various ratios of distances between various facial areas calculated earlier.

Table 1. Various Geometric Features Ratio

S.No	Features
1	Left eye-Nose/right eye-nose
2	Left eye-Mouth/Right eye-Mouth
3	Left eye-Nose/Left eye-Mouth
4	Right eye-Nose/Right eye-Mouth
5	Left eye-Right eye/Nose-Mouth
6	Left eye-Right eye/Left eye-Nose
7	Left eye-Right eye/Left eye-Mouth
8	Right eye-Nose/Nose Height
9	Nose Height/Nose-Mouth

3.1.2Extraction of Skin Color

Different ethnicities have varying skin colors. The dominant skin color of face is used to classify the race as distribution of skin color of different ethnicities is found to be grouped in a small area of color space training set contains both RGB and gray scale images. RGB color space varies largely in intensity. In order to overcome illumination variation, YCbCr color model is adopted and the input image was converted from RGB to YCbCr color space. The conversion is done as follows

$$Cb = -0.148R - 0.291G + 0.439B \tag{1}$$

$$Cr = 0.439R - 0.368G - 0.071B \tag{2}$$

The skin color classification is a good technique to classify ethnicities but it often overlaps for Mongolian and Negroid, though it gives clear picture about a Caucasian.

3.1.3 Normalized forehead area calculation

The normalized forehead area is an important feature for ethnicity classification. Ratio of forehead area and total face area is unique characteristic for face, especially for Negros. Negros usually has large forehead area compared to other races. Forehead area is calculated with the help of Sobel Edge Detection [9] method. The intersection of smoothened vertical and horizontal edges gives the exact eye position. The region above the eye position gives the exact fore head region. Using the coordinates of eyes and the coordinates of face detected earlier, normalized forehead area is calculated as:

$$Normalizedforeheadarea = \frac{Fore\,headarea}{totalfacearea} \tag{3}$$

The forehead to face ratio is found to be more than 25% for Negros and less than 25% for Mongolians. In case of Caucasians, this ratio varies largely

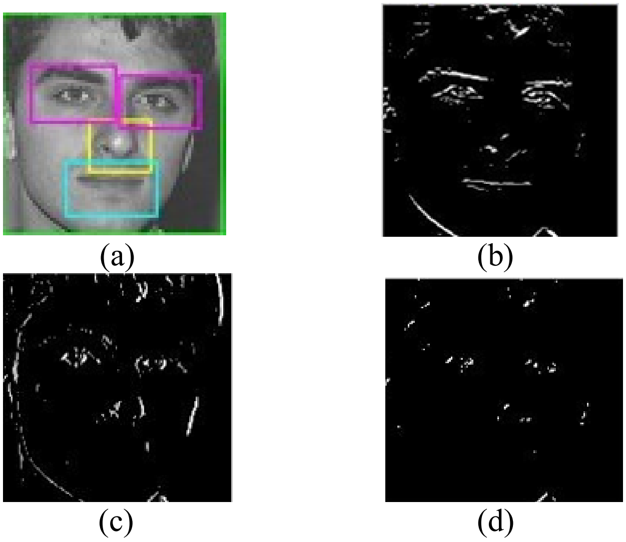

Fig. 3.(a) Detected face, (b) Horizontal Sobel edge (Gx), (c) Vertical Sobel edge (Gy), (d) Intersectoin of Gx and Gy.

3.1.4 Training, Validation and Testing

We implemented a multi-layered perceptrons for the purpose of training. Of the total 447 sample images, 320 images were used training the network 37 for validation and 90 images were used for testing the trained network.

3.2 Experiment 2 (Convolutional Neural Network)

Deep learning has been used for image classification in [10] which has yielded good results. In this experiment, we used a pre-trained model of this 16-layer architecture which has been pre-trained on various faces due to limited computation available with us. This pre-training model as described in [8] was used to extract features from the training and testing samples. During training, input image is passed through the stack of convolution layers which has a filter of size 3*3 and the activation function of "Relu". The VGGNet used has 13 convolution layers. Pooling is done through max pool layer of 2*2 window size and a stride of 2.These stack of convolution layer is followed by three fully connected layer: first two has 4096 channels and the last has 3 channels, each representing a ethnic class.

The network was compiled with categorical crossentropy for loss function (we were using Softmax output) and nesterovmomentum was used. Learning rate was set to 2.5e-4 and freezing all the layers except the final fully connected layers (train at about ~1/10th of the initial learning rate). With all these parameters initialized, this network is used for feature extraction. After this, training and testing was performed.

4. Results

The ethnicity identification model was developed by using 357 images for training, each image being 240*360 in pixels and 90 images for testing purposes. The face samples were taken almost in equal amount: Caucasians 120, Mongolians 120 and Negros 117. Result obtained through ANN and CNN varies largely.

4.1Result of Experiment 1 (ANN)

ANN is trained by changing the various attributes of the network. These networks were then saved and used to evaluate 90 testing samples. The results obtained in various experiments for different classes(C1 C2 and C3) are tabulated below. The best result is achieved with epochs 150, learning rate 0.17, Lavenberg Marquadt training algorithm (trainlm) and hidden neurons 17.

Table 2. Result obtained in various experiments in ANN

ATTRIBUTES	SET1			SET2			SET3		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
Precision	80.6	83.3	83.3	80.6	80.0	83.3	80.0	82.8	78.1
Recall	83.3	80.6	83.3	83.3	77.4	83.3	80.0	77.4	83.3
F-measure	81.9	91.9	83.3	81.9	78.6	83.3	80.0	80.0	80.6
Accuracy	*80.6*	*83.3*	*83.3*	80.6	80.0	83.3	77.4	80.0	83.3

Testing was done using 31 Caucasian, 32 Mongolian and 30 Negroid sample images. The confusion matrix for the experiment 4 shows that out of 31 Caucasians, 25 were classified correctly. In case of Mongolian success rate was 25 out of 31 and in Negroid 25 out of 30.The overall classification accuracy is 82.4% and the misclassification was around is 17.6%.

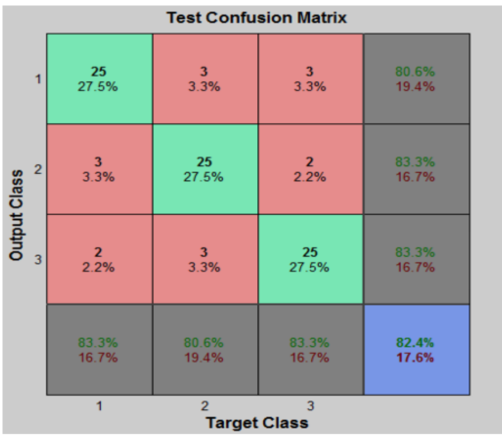

Fig. 4 Confusion Matrix obtained for experiment 1

4.2. Result of Experiment 2 (CNN)

The samples were tested on VGGNet by varying the number of epochs. Two experiments were conducted: one for 10 epochs and the other for 20 epochs. The testing and training accuracies are tabulated below.

Table 3 Results obtained in various experiments in CNN

ATTRIBUTE	Set1	Set2
Epochs	10	20
Training time	413.074sec	824.03sec
Training accuracy	0.9942	0.9967
Testing accuracy	0.93301	0.986

5 Conclusion

Although both the experiments were able to provide the solution to the problem of ethnicity identification among three prominent races: Negroid, Mongolian and Caucasian. But on comparing the results obtained through ANN and CNN, it can be concluded that performance of the CNN approach is far superior than ANN. Accuracy achieved in ANN was 82.4% and in CNN was 98.6%. The results obtained using Artificial Neural Network were better than the results achieved in [4]. And the significant improvement is achieved with convolution neural network.

But the cost of CNN is much more than ANN in terms of time required for feature extraction and training that network. Neural network Training and testing on both the algorithms was done using FERET dataset with 357 training images and 90 testing images. This work could very well be extended for other known ethnicities as well. It can play a major role in surveillance systems for security purposes.

References

1. .Xiaoguang Lu and Anil K. Jain. "Ethnicity identification from face images". Department of Computer Science & Engineering, Michigan State University. In Proceedings of SPIE 5404, Biometric Technology for Human Identification,2004
2. S. Hosoi, E. Takikawa and M. Kawade, "Ethnicity estimation with facial images," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 195-200.
3. G. Guo and G. Mu, "A study of large-scale ethnicity estimation with gender and age variations," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, 2010, pp. 79-86.
4. S. M. M. Roomi, S. L. Virasundarii, S. Selvamegala, S. Jeevanandham and D. Hariharasudhan, "Race Classification Based on Facial Features," *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on*, Hubli, Karnataka, 2011, pp. 54-57.
5. P. Viola and M. Jones. "Robust Real-Time Face Detection". In Proceedings of International

Journal of Computer Vision, May 2004, Volume 57, Issue 2, pp 137–154.
6. FERET Database for facial images. http://user18808:Gz1Bv4Ft@nigos.nist.gov:8080/colorferet/
7. Matlab documentation
8. Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." InProceedings of the British Machine Vision 1.3 (2015).
9. Zhang Jin-Yu, Chen Yan and Huang Xian-Xiang, "Edge detection of images based on improved Sobel operator and genetic algorithms," *2009 International Conference on Image Analysis and Signal Processing*, Taizhou, 2009, pp. 31-35.
10. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, Malaysia, 2015, pp. 730-734.