FACIAL ANIMATION AND MULTIPLE HIERARCHICAL MODELS FOR AGE AND GENDER PREDICT THE FEED FORWARD PROPAGATION AND CASCADE PROPAGATION

Ms Geetha.K, Mr.Aravindhan.S

[*1]M.Phil Research Scholar, Department of Computer science & Applications, Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore District, Tamil Nadu.

Assistant ProfessorDepartment of Computer science & Applications, Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore District, Tamil Nadu .

--***---

ABSTRACT- A person's face provides a lot of information such as age, gender, and identity. Faces play an important role in the estimation/prediction of the age and gender of persons, just by looking at their face. Perceiving human faces and modeling the distinctive features of human faces that contribute most toward face recognition are some of the challenges faced by computer vision and psychophysics researchers. There are many methods have been proposed in the literature for the facial features for age and gender classification. In this research, an attempt is made to classify human age and gender using feed forward propagation neural networks in coarser level. Further final classification is done using 3-sigma control limits in finer level. Proposed approach efficiently classifies three age groups including children, middle-aged adults, and old-aged adults. Similarly two gender groups classified into male and female by the proposed method. The performance of the system is further improved by employing multiple hierarchical decision using three sigma control limits applied on the output of the neural network classifier. The mean and standard deviation has been considered on the output generated from the neural network classifier, and three sigma control limits has been applied to define the range of values for the specific category of age and gender. The efficiency of the system is demonstrated through the experimental results using benchmark database images.

***Key Words:* age and gender prediction, Convention neural network, gaussian mixer model, sigma control**

I. INTRODUCTION

For humanitarian reasons and criminal investigation, the identification of unknown deceased individuals is important. When a collection of bones is discovered, the first thing to ascertain is whether any of the bones are human or of lower animals. This is not easy for a lay-person to distinguish.[1] In particular, difficulties arise where human fetal or newborn skeletal remains are concerned as they often bear little or no resemblance to their adult counterparts and may easily be mistaken as belonging to an animal such as a dog or a rabbit. A trained opinion should always be sought, and if any doubt exists, the bone should be photographed in situ, collected, labeled as to their disposition, and carefully packed and sent for expert laboratory examination

Convolutional neural network (CNN) has achieved promising performance on lots of computer vision and pattern recognition tasks due to its strong ability of self-learning and dealing with large scale data. Although most recent research work focuses mainly on the innovation of CNN structure, the research on label encoding is a development direction of CNN, because a reasonable implementation of label encoding method could help to optimize the CNN structure. Therefore, this thesis mainly explores the implementations of CNN with well-performed label encoding method. In order to evaluate the performance of CNN and label encoding method, two computer vision tasks, age estimation from facial image and depth estimation from a single image, are conducted.-mentioned businesses and probably in many other ones, is named as a fraud detection process. Due to the complexity and enormity of the modern business systems, criminals may and do discover safety gaps and use them to steal data or to defraud somebody. Even if a fraud type is discovered by the authorities and safety regulations are managed, the criminals seek and find other fraudulent ways and thus shift behavior over time. Manual detection conducted by human experts is very expensive even to debug any fraud that has been committed; can't detect all fraudulent transactions of a

certain type; can't be managed to detect the fraudulent behavior the moment it is attempted to be committed and lack the ability to detect the shifts and trends in fraudulent behavior.
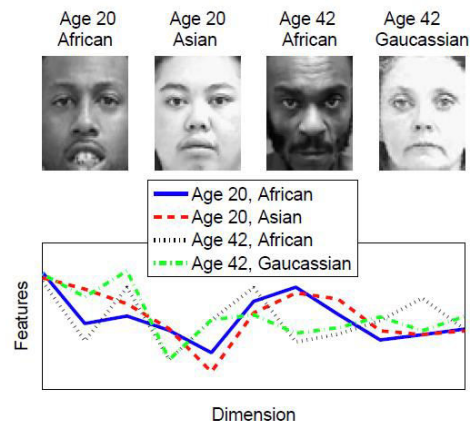
Fig: Facial Image Analysis

sample facial images and their corresponding deep aging features as extracted by our method. Despite of the remarkable intra-class diversity and inter-class similarities in facial appearance of these images, our method can extract broadly similar aging features for similar ages, regardless of race and gender. In the bottom figure, we illustrate the 110-dimension deep aging features for each facial image.

The main objective of this paper we propose a new CNN structure to extract aging features from facial image, which is able to extract both the local and global aging cues. Moreover, we propose a new label encoding method to transfer the discrete aging labels into a continuous possibility vector, which improves the performance of our CNN structure. Our proposed framework achieves state-of-the-art performance on age estimation task.

2. We propose a new framework to conduct depth estimation from a single image task. In this framework, we successfully transfer the regression task into a classification task by applying label encoding method, which improves the performance of our CNN structure by a large margin. Moreover, we implement the surface normal constraints on depth refinement stage, which increases the accuracy of depth estimation. Our proposed framework achieves promising performance on depth estimation from a single image task.

3. Our proposed label encoding method shows a strong capability to improve the performance of CNN without significantly changes of the structure. This could be an interesting research direction of CNN.

This study was done to determine the methods and techniques for the identification of person and estimation of age and sex of the deceased individual. Method: The two problems that arise when human skeletal and dental remains are found are the identification of the person and determination of cause of death. In most cases, establishing the person's identity is not possible and the death of that person remains undetermined. Determination of age sex and identification of deceased person by skeletal bones and by teeth was analyzed. Stature in life, cause of death and Individual Skeletal Characteristics was also analyzed by various means of forensic

procedure. Conclusion: Sex determination by using bone as early as possible is important from the forensic

point of view. The determination of sex by using skull depends upon traits and measurements. The assessment of sex by long bones is easier because the male long bones tend to be longer and more massive than those of the female, with more marked muscle attachments. Teeth are particularly useful in determining gender by using different odontometric techniques.

II.RELATED WORK

A person's face provides a lot of information such as age, gender and identity. Faces play an important role in the estimation/prediction of the age and gender of persons, just by looking at their face. Perceiving human faces and modeling the distinctive features of human faces that contribute most towards face recognition are some of the challenges faced by computer vision and psychophysics researchers. In this research, an attempt is made to classify human age and gender using Multiple Hierarchical decision based on Neural Networks. Now a days, Artificial Neural Network (ANN) has been widely used as a tool for solving many decision modeling problems. In this paper, a feed forward propagation Neural Networks are constructed for human age and gender classification system for gray-scale facial images. Three age groups including Children, Middle-aged adults and Old aged adults and Two gender groups including Male and Female are used in the classification system. The performance of the system is further improved by employing Multiple Hierarchical decision using 3 Sigma Control Limits applied on the output of the Neural Network classifier. The efficiency of the system is demonstrated through the experimental results using benchmark database images[1]. Face detection technique is used for face authentication and verification and face detection is a front part of face recognition. It is used in many fields such as authentication security, video surveillance and human interaction system. In this paper we have collected data of 400 faces from school students in Muzaffarabad, Azad Kashmir. Besides, 50 non-faces are also collected. Both faces and non-faces are preprocessed using Background Elimination, Noise Reduction, Width Normalization and Thinning. After the preprocessing, we have extracted features from 400 faces and 50 non-faces including Geometric Features such as Image Cropping, Vertical/Horizontal Projection, Global Features such as Aspect Ratio, Normalized Area of Faces and Non-faces, Center of Gravity, Slope of Line joining the center of Gravity and texture features. Finally, we have applied Machine Learning Methods such as Bayes, Function, Lazy, Meta, Misc, Rules and Tree to classify the faces and non-faces using 10 fold cross validation. Hyper Pipes gives an overall higher accuracy of 99.8%, while ADTree, LWL and LogiBoost gives accuracy of more than 99%. The average AUC of ROC value was calculated as 96.08%.[2]. In this paper, we describe an automated real-time system that estimates age and gender by utilizing a set of facial image sequences from a video camera. The age and gender estimation system consists of four steps: i) detection and extraction of the facial region from input video; ii) selection of the frontal face images from the extracted facial regions using head pose estimation; iii) duplicated face detection and removal by tracking the faces; and iv) age and gender estimation using statistical facial features. Here, LBP features with AdaBoost classifiers are used to detect the face region in a video frame, and the frontal face images are selected using a 3D pose estimation method. In addition, a particle filter-based tracking framework is employed to remove duplicated faces and to improve the accuracy of people counting, and Gabor-LBP features are used to estimate age

and gender using a linear SVM and Ada boost classifiers. In experiments, a large number of face datasets are used to train and evaluate the proposed method, and higher performance is achieved in terms of age and gender estimation: 72.53% for age and 98.90% for gender.[3]

III. PREVIOUS IMPLEMENTATIONS

Age and gender classification is important visual tasks for human beings, such as many social interactions critically depend on the correct age and gender perception. As visual surveillance and human–computer interaction technologies evolve, computer vision systems for age and gender classification plays an increasing important role in our lives. Age and gender classification is arguably one of the more important visual tasks for an extremely social animal like us humans many social interactions critically depend on the correct gender perception of the parties involved. Arguably, visual information from human faces provides one of the more important sources of information for age and gender classification. Not surprisingly, thus, that a very large number of psychophysical studies has investigated age and gender classification from face perception in humans. Automatic human facial expression recognition, human mood analysis system are the thrust research area in video surveillance and law enforcement applications as a prerequisite for face recognition. Until now much research work has been done on detecting the human faces and recognition of faces but less effort is made for human age and gender prediction. A neural network usually involves a large number of processors operating in parallel, each with its own small sphere of knowledge and access to data in its local memory. Typically, a neural network is initially "trained" or fed large amounts of data and rules about data relationships. A program can then tell the network how to behave in response to an external stimulus or can initiate activity on its own. In feed forward systems, learned relationships about data can "feed forward" to higher layers of knowledge. Neural networks can also learn temporal concepts and have been widely used in signal processing and time series analysis.

Despite recent progress in the area of automatic age prediction from facial images, this area remains a very challenging one for computer vision and pattern recognition.

Important areas of concern include: 1) aging processes are affected by external factors as well as human genetics; 2) males and females may have very different aging characteristics; 3) people of different races have different aging cues. Images of people of the same age may have different facial appearances and images of people of different ages may have similar facial features. Such intra-class diversity and inter-class similarities pose big challenges for automatic age prediction.

3.1 QUANTITATIVE COMPARISON

Here compared our deep feature HADF with the state-of-the-art features for age prediction. Specifically, we extracted the LBP feature and the BIF feature from S3 and S1 and applied the best parameters tested on S2. The LBP feature aims at capturing primarily the skin aging changes and the BIF feature is designed to capture the deep and apparent wrinkles on the face. The dimension of the BIF feature was reduced using PCA from 11080- dimensions to 1000-dimensions to reduce the noise. Recent work has shown that pre-trained CNN features trained on the Image Net can also be transferred to new Classfication or recognition problems

and boost remarkable performance. Here we used the CNN features learnt from ImageNet as aging features for age prediction. trained a CNN for face identification, although it is not for age estimation, their architecture is able to capture high understanding of facial images. Here we used the face identity-preserving features (FIP) for age estimation.
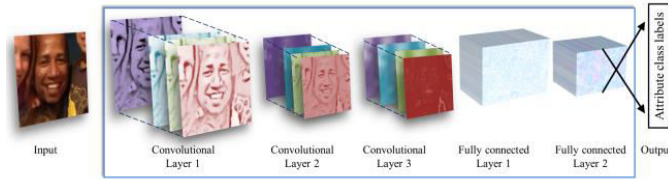

Fig: Illustration of our CNN architecture

Most of the methods discussed above used the FERET benchmark [39] both to develop the proposed systems and to evaluate performances. FERET images were taken under highly controlled condition and are therefore much less challenging than in-the-wild face images. Moreover, the results obtained on this benchmark suggest that it is saturated and not challenging for modern methods. It is therefore difficult to estimate the actual relative benefit of these techniques. As a consequence, [46] experimented on the popular Labeled Faces in the Wild (LFW) [25] benchmark, primarily used for face recognition. Their method is a combination of LBP features with an AdaBoost classifier. As with age estimation, here too, we focus on the Audience set which contains images more challenging than those provided by LFW, reporting performance using a more robust system, designed to better exploit information from massive example training sets.

IV. SYSTEM IMPLEMETNATION

TRAINING AND TESTING

Initialization. The weights in all layers are initialized with random values from a zero mean Gaussian with standard deviation of 0.01. To stress this, we do not use pre-trained models for initializing the network; the network is trained, from scratch, without using any data outside of the images and the labels available by the benchmark. This, again, should be compared with CNN implementations used for face recognition, where hundreds of thousands of images are used for trainingTarget values for training are represented as sparse, binary vectors corresponding to the ground truth classes. For each training image, the target, label vector is in the length of the number of classes (two for gender, eight for the eight age classes of the age classification task), containing 1 in the index of the ground truth and 0 elsewhere.

NETWORK TRAINING.

Aside from our use of a lean network architecture, we apply two additional methods to further limit the risk of over fitting. First we apply dropout learning (i.e. randomly setting the output value of network neurons to zero). The network includes two dropout layers with a dropout ratio of 0.5 (50% chance of setting a neuron's output value to zero). Second, we use data augmentation by taking a random crop of 227×227 pixels from the 256×256 input image and randomly mirror it in each forward-backward training pass. This, similarly to the multiple crop and mirror variations used by [48]. Training itself is performed using stochastic gradient decent

with image batch size of fifty images. The initial learning rate is e −3, reduced to e −4 after 10K iterations.

Prediction. We experimented with two methods of using the network in order to produce age and gender predictions for novel faces:

Center Crop: Feeding the network with the face image, cropped to 227 × 227 around the face center.

Over-sampling: We extract five 227 × 227 pixel crop regions, four from the corners of the 256 × 256 face image, and an additional crop region from the center of the face. The network is presented with all five images, along with their horizontal reflections. Its final prediction is taken to be the average prediction value across all these variations.

We have found that small misalignments in the Audience images, caused by the many challenges of these images (occlusions, motion blur, etc.) can have a noticeable impact on the quality of our results. This second, over-sampling method, is designed to compensate for these small misalignments, bypassing the need for improving alignment quality, but rather directly feeding the network with multiple translated versions of the same face.

ALGORITHMIC IMPLEMENTATION

The filter recurrence relation provides a way to determine the output samples in terms of the input samples and the preceding output. The following pseudo code algorithm will simulate the effect of a high-pass filter on a series of digital samples:

```
// Return RC high-pass filter output samples, given input samples,
// time interval dt, and time constant RC
function high pass (real[0..n] x, real dt, real RC)
  var real[0..n] y
  var real α := RC / (RC + dt)
  y[0] := x[0]
  for i from 1 to n
y[i] := α * y[i-1] + α * (x[i] - x[i-1])
  return y
```

The loop which calculates each of the n {\display style n}   outputs can be refactored into the equivalent:

```
  for i from 1 to n
    y[i] := α * (y[i-1] + x[i] - x[i-1])
```

However, the earlier form shows how the parameter α changes the impact of the prior output y[i-1] and current change in input (x[i] - x[i-1]). In particular

A large α implies that the output will decay very slowly but will also be strongly influenced by even small changes in input. By the relationship between parameter α and time constant R C

{\display style RC} above, a large α corresponds to a large R C {\display style RC} and therefore a low corner frequency of the filter. Hence, this case corresponds to a high-pass filter with a very narrow stop band. Because it is excited by small changes and tends to hold its prior output values for a long time, it can pass relatively low frequencies. However, a constant input (i.e., an input with (x[i] - x[i-1]) =0) will always decay to zero, as would be expected with a high-pass filter with a large R C {\display style RC} .

A small α implies that the output will decay quickly and will require large changes in the input (i.e., (x[i] - x[i-1]) is large) to cause the output to change much. By the relationship between

parameter α and time constant R C {\display style RC} above, a small α corresponds to a

small R C {\display style RC} and therefore a high corner frequency of the filter. Hence, this case corresponds to a high-pass filter with a very wide stop band. Because it requires large (i.e., fast) changes and tends to quickly forget its prior output values, it can only pass relatively high frequencies, as would be expected with a high-pass filter with a small

GAUSSIAN MIXTURE MODEL:

In a Gaussian mixture model, all sets of adaptation requirements are assumed to be sampled from a set of Gaussian distributions spread throughout the space of all possible adaptation requirements. Given a set of training data, the parameters of the distributions are set by running an expectation-maximization (EM) algorithm in order to maximize the likelihood that the given data was sampled from the mixture of distributions. As input for this training procedure, all available history of user adaptation requirements is provided. The number of distributions must be selected a priori, however.

The user has seen and possibly interacted with the images X and A and we must now predict image Y . Based on their required fidelities for X and A, the probability of the user belonging to each Gaussian distribution d, p (d), is computed. This is a calculation over only the images which the user has seen (X and A), in this case given by

$$p(d) = \alpha_d \ \prod_{i=\{X,A\}} \frac{1}{\sigma_d(i) \sqrt{2\pi}} \exp\left(\frac{-(x_i - \mu_d(i))^2}{2\sigma_d(i)^2}\right)$$

For each distribution d with means μd(i), standard deviations σd(i) and prior αd. The distribution with the highest p(d) is selected, call this dbest, and the means of dbest are used to provide a prediction for the adaptation requirements of other images. In this case, since we need to predict Y, we would use μdbest(Y), the mean of the distribution dbest for the image Y.

PRE-PROCESSING:

The main goal of the pre-processing to enhance the visual appearance of images and improve the manipulation of datasets. Pre-processing of image are those operations that are normally required prior to the main data analysis and extraction of information. Image preprocessing, also called image restoration, involves the correction of distortion, degradation,

and noise introduced during the imaging process. Image pre-processing can significantly increase the reliability of an optical inspection. Several filter operations which intensify or reduce certain image details enable an easier or faster evaluation.

Image Adjusting is done with the help of image interpolation. Interpolation is the technique mostly used for tasks such as zooming, rotating, shrinking, and for geometric corrections. Removing the noise is an important step when image processing is being performed. However noise may affect

segmentation and pattern matching. When performing smoothing process on a pixel, the neighbor of the pixel is used to do some transforming. After that a new value of the pixel is created.

GRAY SCALE CONVERSION:

The image acquired is in RGB color. It is converted into gray scale because it carries only the intensity information which is easy to process instead of processing three components R (Red), G (Green), B (Blue). to take the RGB values for each pixel and make as output a single value reflecting the brightness of that pixel. One such approach is to take the average of the contribution from each channel: $(R+B+C)/3$. However, since the perceived brightness is often dominated by the green component, a different, more "human oriented", method is to take a weighted average, e.g.: $0.3R + 0.59G + 0.11B$.

EDGE DETECTION:

Edge detection is the name for a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. The points at which image brightness changes sharply are typically organized into a set of curved line segments termed edges. The same problem of finding discontinuities in 1D signals is known as step detection and the problem of finding signal discontinuities over time is known as change detection. Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction. Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as image processing, computer vision, and machine vision.

Based on this one-dimensional analysis, the theory can be carried over to two dimensions as long as there is an accurate approximation to calculate the derivative of a two-dimensional image. The Sobel operator performs a 2-D spatial gradient measurement on an image. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. The Sobel edge detector uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows). A convolution mask is usually much smaller than the actual image. As a result, the mask is slid over the image, manipulating a square of pixels at a time. The actual Sobel masks are shown below:

The magnitude of the gradient is then calculated using the formula:

$$|G| = \sqrt{Gx^2 + Gy^2}$$

An approximate magnitude can be calculated using:

$$|G| = |Gx| + |Gy|$$

The code for the Sobel edge detector is shown below and uses the above gradient approximation.

IMAGE SEGMENTATION:

Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. Image restoration is different from image enhancement in that the latter is designed to emphasize features of the image that make the image more pleasing to the observer, but not necessarily to produce realistic data from a scientific point of view. Image enhancement techniques (like contrast stretching or de-blurring by a nearest neighbor procedure) provided by "Imaging packages" use no a priori model of the process that created the image. With image enhancement noise can effectively be removed by sacrificing some resolution, but this is not acceptable in many applications. In a Fluorescence Microscope resolution in the z-direction is bad as it is. More advanced image processing techniques must be applied to recover the object. Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

FEATURE EXTRACTION:

Feature extraction is a special form of dimensional reduction. When the input data to an algorithm is too large to be processed and it is suspected to be very redundant then the input data will be transformed into a reduced representation set of features. Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

EXPERIMENTAL RESULT

Our method is implemented using the Cafe open-source framework [14]. Training was performed on an Amazon GPU machine with 1,536 CUDA cores and 4GB of video memory. Training each network required about four hours, predicting age or gender on a single image using our network requires about 200ms. Prediction running times can conceivably be substantially improved by running the network on image batches.

Fig: Gender misclassifications

Evidently, the proposed method outperforms the reported state-of-the-art on both tasks with considerable gaps. Also evident is the contribution of the over-sampling approach, which provides an additional performance boost over the original network.

Fig: Age misclassifications

We provide a few examples of both gender and age misclassifications in Figures 4 and 5, respectively. These show that many of the mistakes made by our system are due to extremely challenging viewing conditions of some of the Audience benchmark images. Most notable are mistakes caused by blur or low resolution and occlusions (particularly from heavy makeup). Gender estimation mistakes also frequently occur for images of babies or very young children where obvious gender attributes are not yet visible.
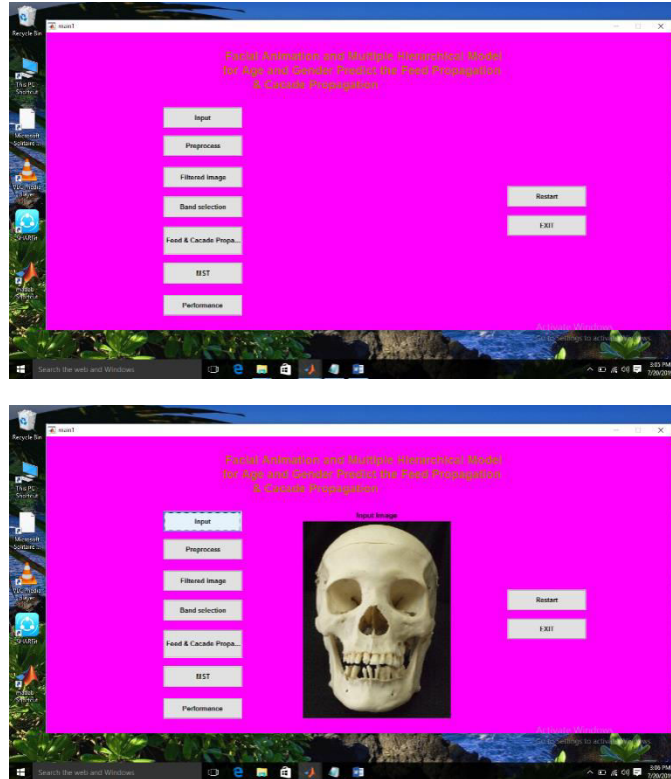
Fig: input Image

In this research, there are 1000 gray scale with 256 gray levels facial images used for experiment. Each image size is normalized to 64×64. Among the 1000 experimental images, 700 images are used as the training data and the remaining are used as test images. The male face images are trained using to neural networks, and core Hierarchical classification is done using 3 Sigma limits.
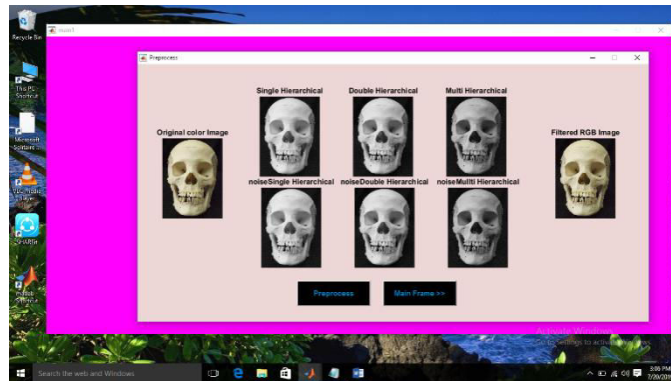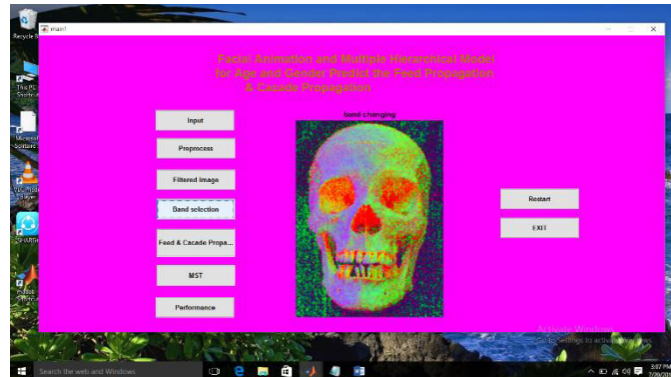

Fig: preprocessing

In the testing phase, among 300 images the combination of male and female of different age groups taken. Thus, the success rate for male and female images different expressions are 94.00% and 96.00 for respectively. Therefore, the overall success rate for images is 95.00%. The

average recognition time of each test images for 0.30 seconds on 2GBa Pentium Quad Core proc.RAM.

The proposed method can be detect the effect in medical imaging, sensing any kind of surveillance systems. However, proposed method fail side-view faces, occluded faces and partial face images. This is due to the fact that the proposed model is a constrained to detect only the frontal view face. Our proposed method is compared with the automatic detect the face and age gender to the Age Estimation Based on Facial Aging and gender to prediction
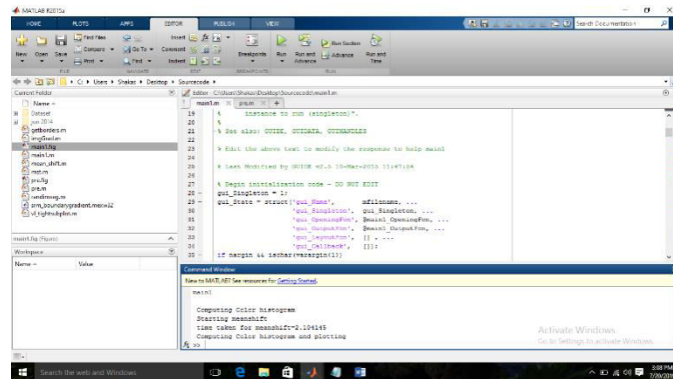

Fig: Accuracy prediction

It provides a confusion matrix for our multi-class age classification results. For age classification, we measure and compare both the accuracy when the algorithm gives the exact age-group classification and when the algorithm is off by one adjacent age-group (i.e., the subject belongs to the group immediately older or immediately younger than the predicted group). This follows others who have done so in the past, and reflects the uncertainty inherent to the task – facial features often change very little between oldest faces .
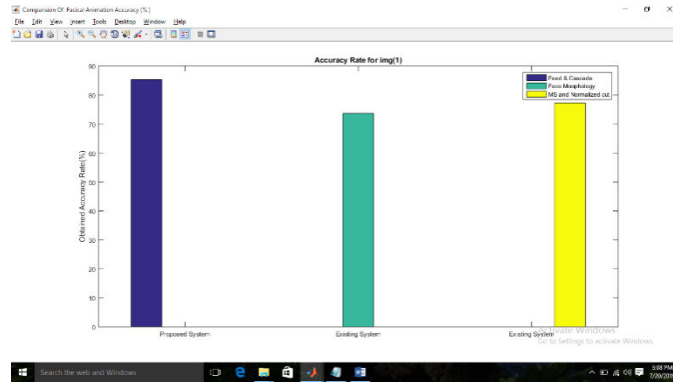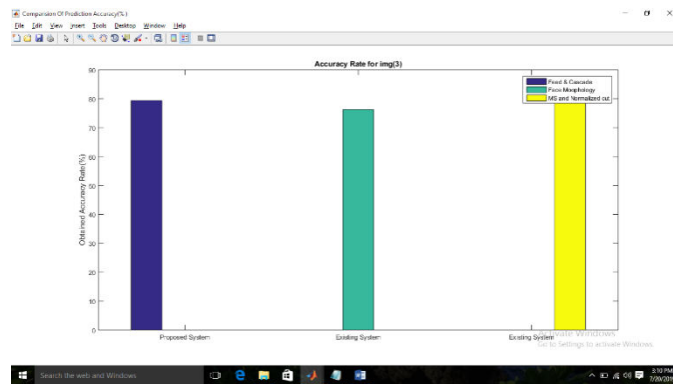
Fig: Accuracy Rate for Age

Fig: Accuracy Rate Gender

CONCLUSION

Though many previous methods have addressed the problems of age and gender classification, until recently, much of this work has focused on constrained images taken in lab settings. Such settings do not adequately reflect appearance variations common to the real-world images in social websites and online repositories. Internet images, however, are not simply more challenging: they are also abundant. The easy availability of huge image collections provides modern machine learning based systems with effectively endless training data, though this data is not always suitably labeled for supervised learning. Taking example from the related problem of face recognition we explore how well deep CNN perform on these tasks using Internet data. We provide results with a lean deep-learning architecture designed to avoid over fitting due to the limitation of limited labeled data. Our network is "shallow" compared to some of the recent network architectures, thereby reducing the number of its parameters and the chance for over fitting. We further inflate the size of the training data by artificially adding cropped versions of

the images in our training set. The resulting system was tested on the Audience benchmark of unfiltered images and shown to significantly outperform recent state of the art. Two important conclusions can be made from our results. First, CNN can be used to provide improved age and gender classification results, even considering the much smaller size of contemporary unconstrained image sets labeled for age and gender. Second, the simplicity of our model implies that more elaborate systems using more training data may well be capable of substantially improving results beyond those reported here.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. Trans. Pattern Anal. Mach. Intel., 28(12):2037–2041, 2006.

[2] S. Baluja and H. A. Rowley. Boosting sex identification performance. Int. J. Comput. Vision, 71(1):111–119, 2007.

[3] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In Int. Conf. Mach. Learning, volume 3, pages 11–18, 2003.

[4] W.-L. Chao, J.-Z. Liu, and J.-J. Ding. Facial age estimation based on label-sensitive learning and age-oriented regression. Pattern Recognition, 46(3):628–641, 2013. 1,

[5] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531, 2014. 3

[6] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: A robust local image descriptor. Trans. Pattern Anal. Mach. Intel., 32(9):1705–1720, 2010.

[7] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim. Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recognition, 44(6):1262–1281, 2011.

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In European Conf. Comput. Vision, pages 484–498. Springer, 1998.

[9] C. Cortes and V. Vapnik. Support-vector networks. Machine learning, 20(3):273–297, 1995.

[10] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. Trans. on Inform. Forensics and Security, 9(12), 2014. 1, 2, 5,

[11] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. Trans. Pattern Anal. Mach. Intel., 32(11):1955–1976, 2010.

[12] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. Int. Conf. Multimedia, 10(4):578–584, 2008.

[13] K. Fukunaga. Introduction to statistical pattern recognition. Academic press, 1991.

[14] A. C. Gallagher and T. Chen. Understanding images of groups of people. In Proc. Conf. Comput. Vision Pattern Recognition, pages 256–263. IEEE, 2009. 2,

[15] F. Gao and H. Ai. Face age classification on consumer images with gabor feature and fuzzy lda method. In Advances in biometrics, pages 132–141. Springer, 2009. 1,

[16] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. Trans. Pattern Anal. Mach. Intel., 29(12):2234–2240, 2007. 2

[17] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In Neural Inform. Process. Syst., pages 572–579, 1990.

[18] A. Graves, A.-R. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pages 6645–6649. IEEE, 2013.

[19] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang. Image based human age estimation by manifold learning and locally adjusted robust regression. Trans. Image Processing, 17(7):1178–1188, 2008. 2