ISSN 2047-3338

Feature Extraction from Human Facial Images

Arati Jhunjhunwala[1] and Samir Kumar Bandyopadhyay[2]
[1,2]Department of Computer Science and Engineering, University of Calcutta, India
[1]aratigrwl6@gmail.com, [2]skb1@vsnl.com

***Abstract*– Due to technological advancements there is an arousal of the world where human being and intelligent robots live together. Area of Human Computer Interaction (HCI) plays an important role in resolving the absences of neutral sympathy in interaction between human being and machine (computer). HCI will be much more effective and useful if computer can predict about emotional state of human being and hence mood of a person from supplied images on the basis of facial expressions. It is observed that 7% of human communication information is communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. For classifying facial expressions into different categories, it is necessary to extract important facial features which contribute in identifying proper and particular expressions. Recognition and classification of human facial expression by computer is an important issue to develop automatic facial expression recognition system in vision community. This paper introduces some novel models for all steps of a face recognition system from a non-frontal RGB image. For face portion segmentation (face detection) and localization, morphological image processing operations are used. Permanent facial features like eyebrows, eyes, mouth and nose are extracted using SUSAN edge detection operator, facial geometry, edge projection analysis.**

***Index Terms*– HCI, Face Detection, Non-Frontal Face, RGB Image Facial Features, and Segmentation SUSAN Edge Detection Operator**

I. INTRODUCTION

THE analysis of the human face via image (and video) is one of the most interesting and focusing research topics in the last years for the image community. From the analysis (sensing, computing, and perception) of face images, much information can be extracted, such as the sex/gender, age, facial expression, emotion/temper, mentality/mental processes and behavior/psychology, and the health of the person captured. According to this information, many practical tasks can be performed and completed; these include not only person identification or verification (face recognition), but also the estimation and/or determination of person's profession, hobby, name (recovered from memory), etc.

Research on face image analysis has been carried out and is being conducted around various application topics, such as (in alphabetical order) age estimation, biometrics, biomedical instrumentations, emotion assessment, face recognition, facial expression classification, gender determination, human-computer/human-machine interaction, human behavior and emotion study, industrial automation, military service, psychosis judgment, security checking systems, social signal processing, surveillance systems, sport training, tele-medicine service, etc.

Therefore facial expressions are the most important information for emotions perception in face to face communication. This paper explains about an approach to the problem of facial feature extraction from a non-l frontal posed image For face portion segmentation basic image processing operation like morphological dilation, erosion, reconstruction techniques with disk structuring element are used. Six permanent Facial features like eyebrows(left and right), eye (left and right) , mouth and nose are extracted using facial geometry, edge projection analysis and distance measure and feature vector is formed considering height and width of left eye, height and width of left eyebrow, height and width of right eye, height and width of right eyebrow, height and width of nose and height and width of mouth along with distance between left eye and eyebrow, distance between right eye and eyebrow and distance between nose and mouth.

A) Image Enhancement

In RGB images each pixel has a particular color; that color is described by the amount of red, green and blue in it. If each of these components has a range 0–255, this gives a total of 256^3 different possible colors. Such an image is a "stack" of three matrices; representing the red, green and blue values for each pixel. This means that for every pixel there correspond 3 values. Whereas, in greyscale each pixel is a shade of gray, normally from 0 (black) to 255 (white). This range means that each pixel can be represented by eight bits, or exactly one byte. Other greyscale ranges are used, but generally they are a power of 2. So, we can say gray image takes less space in memory in comparison to RGB images.

B) Face Detection

In this paper, we focus on only machine learning methods because they eliminate subjective thinking factors from human experience. Moreover, they only depend on training

data to make final decisions. Thus, if training data is well organized and adequate, then these systems will achieve high performance without human factors.

One of the most popular and efficient learning machine based approaches for detecting faces is AdaBoost approach. Viola and Jones designed a fast, robust face detection system where AdaBoost learning is used to build nonlinear classifiers. AdaBoost is used to solve the following three fundamental problems: (1) learning effective features from a large feature set; (2) constructing weak classifiers, each of which is based on one of the selected features; (3) boosting the weak classifiers to construct a strong classifier.

One of the popular methods having the same achievement as well is artificial neural networks (ANNs) [7]. ANN is the term on the method to solve problems by simulating neuron's activities. In detail, ANNs can be most adequately characterized as "computational models" with particular properties such as the ability to adapt or learn, to generalize, or to cluster or organize data, and which operation is based on parallel processing.

Our hybrid model is named ABANN. This is the model of combining AB and ANN for detecting faces. In this model, ABs have a role to quickly reject nonface images; then ANNs continue filtering false negative images to achieve better results. The final result is face/nonface. The selected neural network here is three-layer feed forward neural network with back propagation algorithm. The number of input neurons T is equivalent to the length of extracted feature vector, and the number of output neurons is just 1 (C = 1), This will return *true* if the image contains a human face and *false* if it does not.

The number of hidden neurons H will be selected based on the experiment; it depends on the sample database set of images. The output of the ANN is a real value between −1 (false) and +1 (true).

C) Face Alignment

The face alignment is one of the important stages of the face recognition. Moreover, face alignment is also used for other face processing applications, such as face modeling and synthesis. Its objective is to localize the feature points on face images such as the contour points of eye, nose, mouth, and face.

The method proposed in this paper is similar to the one of Rowley *et al*. (Rowley et al., 1998) but it not only corrects in-plane rotation but also *x*/*y* translation and scale variations. It is further capable of treating *non-frontal* face images and employs an iterative estimation approach. The system makes use of a Convolutional Neural Network (CNN) (LeCun, 1989; Le Cun et al., 1990) that, after being trained, receives a mis-aligned face image and directly and simultaneously responds with the respective parameters of the transformation that the input image has undergone.

The proposed neural architecture is a specific type of neural network consisting of seven layers, where the first layer is the input layer, the four following layers are convolutional and sub-sampling layers and the last two layers are standard feed-forward neuron layers. The aim of the system is to learn a function that transforms an input pattern representing a mis-aligned face image into the four transformation parameters corresponding to the misalignment, *i.e*., x/y translation, rotation angle and scale factor:

1) The retina $l1$ receives a cropped face image of 46×56 pixels, containing gray values normalized between −1 and +1. No further pre-processing like contrast enhancement, noise reduction or any other kind of filtering is performed.
2) The second layer $l2$ consists of four so-called feature maps. Each unit of a feature map receives its input from a set of neighboring units of the retina This set of neighboring units is often referred to as *local receptive field*, a concept which is inspired by Hubel and Wiesel's discovery of locally-sensitive, orientation-selective neurons in the cat visual system (Hubel and Wiesel, 1962). Such local connections have been used many times in neural models of visual learning (Fukushima, 1975; LeCun, 1989; LeCun et al., 1990; Mozer, 1991). They allow extracting elementary visual features such as oriented edges or corners which are then combined by subsequent layers in order to detect higher-order features. Clearly, the position of particular visual features can vary considerably in the input image because of distortions or shifts. Additionally, an elementary feature detector can be useful in several parts of the image. For this reason, each unit shares its weights with all other units of the same feature map so that each map has a fixed feature detector. Thus, each feature map $y2i$ of layer $l2$ is obtained by convolving the input map $y1$ with a trainable kernel $w2i$:

   $y2i(x,y) = \text{å}\ (u,v)\in K$

   $w2i(u,v)y1(x+u,y+v) + b2i$, where $K = \{(u,v) \mid 0 < u < sx\ ;\ 0 < v < sy\}$ and $b2i \in$ R is a trainable *bias* which compensates for lighting variations in the input. The four feature maps of the second layer performeach a different 7×7 convolution ($sx = sy = 7$). Note that the size of the obtained convolutional maps in $l2$ is smaller than their input map in $l1$ in order to avoid border effects in the convolution.
3) Layer $l3$ sub-samples its input feature maps into maps of reduced size by locally summing up the out- put of neighboring units.Further, this sum is multiplied by a trainable weight $w3\ j$, and a trainable bias $b3\ j$ is added before applying a sigmoid activation function $F(x) = \arctan(x)$: $y3\ j(x,y) = F\_w3\ j\ \text{å}\ u,\ v \in \{0,1\}\ y2\ j(2x+u,2y+v)+b3\ j$.

   Thus, sub-sampling layers performsome kind of averaging operation with trainable parameters. Their goal is to make the system less sensitive to small shifts, distortions and variations in scale and rotation of the input at the cost of some precision.
4) Layer $l4$ is another convolutional layer and consists of three feature maps, each connected to two maps of the preceding layer $l3$. In this layer, 5×5 convolution kernels are used and each featuremap has two different convolution kernels, one for each input map. The results of the two convolutions as well as the bias are simply added up. The goal of layer $l4$ is to extract higher-level features by combining lower level information from the preceding layer.
5) Layer $l5$ is again a sub-sampling layer that works in the same way as $l3$ and again reduces the dimension of the

respective feature maps by a factor two. Whereas the previous layers act principally as feature extraction layers.

6) Layers *l*6 and *l*7 combine the extracted local features from layer *l*4 into a global model. They are neuron layers that are fully connected to their respective preceding layers and use a sigmoid activation function. *l*7 is the output layer containing exactly four neurons, representing x and y translation, rotation angle and scale factor, normalized between −1 and +1.

After activation of the network, these neurons contain the estimated normalized transformation parameters *y*7*i* of the mis-aligned face image presented at *l*1. Each final transformation parameter *pi* is then calculated by linearly rescaling the corresponding value *y*7*i* from [-1,+1] to the interval of the respective minimal and maximal allowed values *pmini* and *pmaxi*:*pi* =(*pmaxi*− *pmini)/2*(*y*7*i*+1)+ *pmini* , *i* = 1..4

Using the annotated facial features, we were able to crop well aligned face images where the eyes and the mouth are roughly at pre-defined positions in the image while keeping a constant aspect ratio. By applying transformations on the well-aligned face images, we produced a set of artificially mis-aligned face images that we cropped from the original image and resized to have the dimensions of the retina (*i.e*., 46×56). The transformations were applied by varying the translation between −6 and +6 pixels, the rotation angle between −30 and +30 degrees and the scale factor between 0.9 and 1.1.

The respective transformation parameters *pi* were stored for each training example and used to form the corresponding desired outputs of the neural network by normalizing them between −1 and +1:

di =2(*pi*− *pmini*)/(*pmaxi*− *pmini)*-1

The objective function is simply the Mean Squared Error (MSE) between the computed outputs and the desired outputs of the four neurons in *l*7. At each iteration, a set of 1,000 face images is selected at random. Then, each face image example of this set and its known transformation.

Classically, in order to avoid overfitting, after each training iteration, a validation phase is performed using a separate validation set. A minimal error on the validation set is supposed to give the best generalization, and the corresponding weight configuration is stored. A correction of the bounding box can then simply be achieved by applying the inverse transformation parameters (−*pi* for translation and rotation, and 1/*pi* for scale). However, in order to improve the correction, this step is performed several (*e.g*. 30) times in an iterative manner, where at each iteration, only a certain proportion (*e.g*., 10%) of the correction is applied to the bounding box. Then, the face image is re-cropped using the new bounding box and a new estimation of the parameters is calculated with this modified image. The transformation with respect to the initial bounding box is obtained by simply accumulating the respective parameter values ateach iteration. Using this iterative approach, the system finally converges to a more precise solution than when using a full one-step correction. Moreover, oscillations can occur during the alignment cycles. Hence, the solution can further be improved by reverting to that iteration where the neural network estimates the minimal transformation, *i.e*., where the outputs *y*7*i* are the closest to zero.

D) Facial Feature Extraction

Fig. 1: Facial feature extraction

Face area and facial feature plays an important role in facial expression recognition**.** Better the feature extraction rate more is the accuracy of facial expression recognition. Precise localization of the face plays an important role in feature extraction, and expression recognition. But in actual application, because of the difference in facial shape and the quality of the image, it is difficult to locate the facial feature precisely (Fig. 1).

In order to extract facial features, segmented face image (RoI) is then resized to larger size to make facial components more prominent. SUSAN edge detection operator along with noise filtering operation is applied to locate the edges of various face feature segment components. SUSAN operator places a circular mask around the pixel in question. It then calculates the number of pixels within the circular mask which have similar brightness to the nucleus and refers it as USAN and then subtract USAN size from geometric threshold to produce edge strength image.

Following steps are utilized for facial feature segment localization:

1. Apply Morphological operations to remove smaller segments having all connected components (objects) that have fewer than P pixels where P is some threshold value.
2. Trace the exterior boundaries of segments and draw bounding box by taking into account x,y coordinates and height and width of each segment.
3. Image is partitioned into two regions i.e upper and lower portion on the basis of centre of cropped face. Assuming the fact that eyes and eyebrows are present in the upper part of face and mouth and nose is present in the lower part. Smaller segments within the region are eliminated by applying appropriate threshold value and remaining number of segments are stored as upper left index, upper right index and lower index.

   Following criteria is used for selecting appropriate upper left index, upper right index and lower index.

   a) A portion is an upper part if x and y values are less than centx and centy where centx and centy are x- and ycoordinates of center of cropped image. Eyes and eyebrows are present in this area. For left eye and eyebrow portion certain threshold for values of x and y is considered for eliminating outer segments. For right

eye and eyebrow also specific threshold value is chosen for eliminating outer segments.

b) A portion is a lower portion if its value is greater than centx and centy where centx and centy are x- and y coordinates of center of an image. Nose and mouth are present in this area. For nose and mouth area segments, x lies in the specific range and y also lies in certain range is considered as region of interest for eliminating outer segments .Here number of segments for each portion are stored. If number of segments are > 2 then following procedure for combining the segments is called

4. Segments are checked in vertical direction. If there is overlapping then the segments are combined. Again if segments are >2 then distance is obtained and the segments which are closer are combined. This process is repeated until we get two segments for each part and in all total six segments This gives the bounding box for total six segments which will be left and right eyes , left and right eyebrows , nose and mouth features of the supplied face.

E) Formation of Feature Vector

Bounding box location of feature segments obtained in the above step are used to calculate the height and width of left eyebrow, height and width of left eye, height and width of right eyebrow, height and width of right eye, height and width of nose and height and width of mouth. Distance between centre of left eye and eyebrow, right eye and eyebrow and mouth and nose is also calculated. Thus total 15 parameters are obtained and considered as feature vector:

Thus,

Fv={H1,W1,H2,W2,H3,W3,H4,W4,Hn,Wn,Hm,Wm,D1,D2, D3}

Where,

H1=height of left eyebrow, W1= width of left eyebrow

H2= height of left eye, W2= width of left eye

H3=height of right eyebrow, W3= width of right eyebrow

H4= height of right eye, W4= width of right eye

Hn= height of nose, Wn= width of nose,

Hm= height of mouth, Wm= width of mouth

D1 = distance between centre of left eyebrow and left eye,

D2= distance between centre of right eyebrow and right eye,

D3= distance between centre of nose and mouth

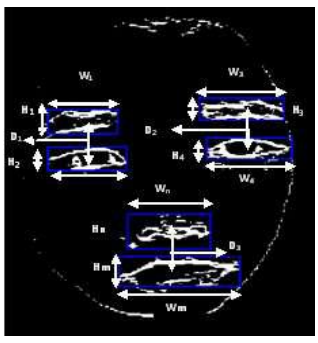

Fig. 1: Formation of Feature Vector

II. METHODOLOGY FOR OUR PROPOSED WORK

1. Read input image
2. Image enhancement :Converting RGB image to Gray Scale
3. Face Detection using Adaptive Boosting and Artificial Neural Network
4. Face Alignment using Convolution Neural Network (CNN)
5. Crop the Aligned Face Image
6. Resize
7. Extract features from cropped face

The principle of operator SUSAN is to make a mask on the circle area of one point with the radius of *r*. *t* is the threshold of difference between pixels. Beyond it, two pixels will be taken for having different luminance. The *g* is the geometry threshold, which means the max value of the USAN area. Beyond it, that pixel is not on the edge.

According to the properties of operator SUSAN, it can be used not only to detect the edge, but also to extract the corner point. Therefore, comparing with the operator such as Sobel, Canny, and so on, it is more appropriate to extract the features of eyes and mouth on the face, especially to locate the corner points of eyes and mouth automatically. It also can get good achievements on images with various qualities by adjusting the parameters *r*, t and of g operator SUSAN.

III. SURVEY OF EXISTING METHODS

Facial feature contains three types of information i.e., texture, shape and combination of texture and shape information. Feng et al. [15] used LBP and AAM for finding combination of local feature information, global information and shape information to form a feature vector. They have used nearest neighborhood with weighted chi-sq statistics for expression classification. Feature point localization is done using AAM and centre of eyes and mouth is calculated based on them.

Mauricio Hess and G. Martinez [16] used SUSAN algorithm Comparison of the recognition performance with different types of features shows that Gabor wavelet coefficients are more powerful than geometric positions. Junhua Li and Li Peng [24] used feature difference matrix for feature extraction and QNN (Quantum Neural Network) for expression recognition from the survey, it is observed that various approaches have been used to detect facial features [25] and classified as holistic and feature based methods to extract facial feature from images or video sequences of faces. These are geometry based, appearance based, template based and skin color segmentation based approaches. Recently large amount of contributions were proposed in recognizing expressions using dynamic textures features using both LBP and gabor wavelet approach and appearance features and increases complexity. Moreover one cannot show features located with the help of bounding box. Hence, the proposed facial expression recognition system aimed to use image

preprocessing and geometry based techniques for feature extraction.

IV. CONCLUSION AND FUTURE SCOPE

The enhancement of thermal images is useful in quality control, problem diagnostics, research and development. The image obtained after the morphology operation is much clear than the other images, that will be helpful in problem diagnostics. This paper presented novel models for all steps of the recognition of human faces in 2-dimentional digital images. The results show that ABANN not only gets approximate detection rate and processing time AdaBoost detector but also minimizes false detections. ABANN had solved the drawbacks of AdaBoost and ANN detector. a novel technique that aligns faces using their respective bounding boxes coming from a face detector.

The method is based on a convolutional neural network that is trained to simultaneously output the transformation parameters corresponding to a given misaligned face image. In an iterative and hierarchical approach this parameter estimation is gradually refined. The system is able to correct translations of up to 13% of the face bounding box width, inplane rotations of up to 30 degrees and variations in scale from 90% to 110%. The combination of SUSAN edge detector, edge projection analysis and facial geometry distance measure is best combination to locate and extract the facial feature for gray scale images in constrained environments Therefore in future an attempt can be made to develop hybrid approach for facial feature extraction and recognition accuracy can be further improved using same NN approach and hybrid approach such as ANFIS. An attempt can also be made for recognition of other database images or images captured from camera.

REFERENCES

[1] Thai Hoang Le, Department of Computer Science, Ho Chi Minh University of Science, Ho Chi Minh City 70000, Vietnam, 2011, Applying Artificial Neural Networks for Face recognition.

[2] Stefan Duffner and Christophe Garcia, Robust Face Alignment Using Convolutional Neural Networks.

[3] S.P. Khandait, R.C. Thool, P.D. Khandait, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.1, January 2011, Automatic Facial Feature Extraction and Expression Recognition based on Neural Network.

[4] Hua Gu Guangda Su Cheng Du,Research Institute of Image and Graphics, Department of Electronic Engineering, Tsinghua University, Beijing, “Feature Points Extraction from Faces”.