Face Detection in Skin-Toned Images Through Wavelet Edges and Neural Network

H. C. Vijay Lakshmi and Sudarshan PatilKulkarni

Abstract—**In this paper, an improved algorithm to detect faces in images with skin tone regions is proposed. Among the segmented candidate regions, facial edges detected using Canny and Sobel operators on wavelet approximation are suggested as feature set. These features are further classified using a neural-net proposed. Results of the test using the proposed algorithm are compared with those of the previous ones.**

Index Terms—**Face detection, localization, edge, segment, approximation.**

I. INTRODUCTION

Face detection and localization is the task of checking whether the given input image or video sequence contains any human face, and if human faces are present, returning the location of the human face in the image. The faces play a major role in identifying and recognizing people. The area of face detection has gained considerable importance with the advancement of human-machine interaction as it provides a natural and efficient way to communicate between humans and machines. Face detection and localization in image sequences has become a popular area of research due to emerging applications in intelligent human-computer interface, surveillance systems, content-based image retrieval, video conferencing, financial transaction, forensic applications, and many other fields. Face detection is essentially localizing and extracting a face region from the background. This may seem like an easy task but the human face is a dynamic object and has a high degree of variability in its appearance [1], which makes face detection a difficultproblem in computer vision. A large number of factors that govern the problem of face detection [2], [3].The long list of these factors include the pose, orientation, facial expressions, facial sizes found in the image, luminance conditions, occlusion, structural components, gender, ethnicity of the subject, the scene and complexity of image's background. Faces appear totally different under different lighting conditions. A thorough survey of face detection research work is available in [2], [3]. In terms of applications, face detection and good localization is an important preprocessing step in online face recognition A number of techniques have been developed by researchers in order to efficiently detect human faces in any given input image or video sequence.

Segmentation of the input image based on skin

chromaticity is the first step in detecting and localizing faces in color images. Segmentation of an image based on human skin chromaticity using different color spaces results in identifying even pseudo skin like regions as skin regions. Hence there is a need for further eliminating these pseudo skin regions. Researchers are working on adaptive skin color segmentation used for detection [4]. The segmentation using combination of color spaces combined with Canny and Prewitt edge detection for obtaining the region boundaries segment better when compared with the combination of YCbCr color space and Robert Cross edge [5], [6]. There are several challenges while detecting and locating faces in skin toned regions, refer [6]. Inspite of using combination of different color spaces during segmentation, it is tedious to demarcate region boundaries between skin and pseudo skin regions and also eliminate these regions from searching process "Fig. 1a" and "Fig. 1b". The use of color space alone sometimes fails to segment the boundary regions of the image. In order to overcome this problem combination of colour spaces for efficient skin segmentation followed by Canny and Prewitt edge detection to demarcate the region boundary is used for input image segmentation [6]. Due to variation in illumination, skin regions may not be properly identified as skin during skin segmentation. Locating faces in these circumstances is more complex as opposed to localizing faces with uniform, non skin-tone background. The "Open CV" face detection [7] software based on Viola-Jones [8] correctly identifies the frontal faces in the images, but fails to the dataset containing multiple faces with skintone background and dresses, refer "Fig. 1c"

II. EDGE DETECTION

Edge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. In the context of the problem under consideration it is the facial boundaries that need to be identified correctly. In the proposed algorithm, use of Canny and Sobel methods are suggested for facial edge extraction. The Canny method finds edges by looking for local maxima of the gradient of image. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to respond to noise, and more likely to detect true weak edges. The Canny method applies a high threshold for low edge sensitivity and a low threshold for high edge sensitivity. Edge starts with the low sensitivity result and then grows it to include connected edge pixels

from the high sensitivity result. This helps fill in gaps in the detected edges. Canny's edge detection algorithm performs better than other methods under almost all scenarios and performs well even under noisy conditions [9].

Fig. 1a. Input image   Fig. 1b. Segmented image

Fig. 1c. Detected faces using open CV

Face detection is a classification problem; it calls for classifying the selected segmented region as face or non-face. For efficient classification a robust feature set as well as robust classification method is very much essential. This calls for an efficient feature extraction method followed by efficient classification method. Edges play an important role in extracting prominent facial features. For detailed analysis of various edge detection algorithms refer [5], [9]. Two level wavelet decomposition [10] of face images followed by edge extraction for approximation image results in prominent facial edges, which is evident from the images of the facial edges extracted for raw face image "Fig. 2a" and "Fig. 2b", edges extracted for the approximation images after first and second level wavelet decomposition "Fig. 2c" and "Fig. 2d" respectively. The edges extracted for the second level approximation eliminates all unwanted noisy edges present in the image and also reduces the size of the edge feature set due to two level wavelet decomposition of the input image.

Fig. 2a. Input face image   Fig. 2b. Edge extracted raw input image

Fig. 2c. Edge extraction of   Fig. 2d. Edge extraction offirst level
approximation image   second level approximation image

III. NEURAL NETWORK

Neural network is a machine learning approach functions the way in which brain performs a particular learning task. Artificial neural network finds its application while solving problems involving pattern classification, clustering or categorization, function approximation, optimization,

content addressable memory etc. A neuron is information processing unit that is fundamental to the operation of a neural network and it s as shown in the "Fig 3".

Artificial neural network can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges are connections between neuron outputs and neuron inputs. A learning rule is defined as a procedure for modifying the weights and biases. Knowledge about the

Learning task is given in the form of examples. Inter neuron connection strength (weights) are used to store the acquired information through training. During training process the weights are modified in order to model the particular task correctly on the basis of training examples.

Learning can be classified as supervised learning and unsupervised learning. In supervised learning the network is provided with a correct answer for every input pattern. Weights are determined to allow the network to produce answers as close as possible to the known correct answers. In unsupervised learning, it does not require a correct answer associated with each input pattern in the training data set. It explores the underlying structure in the data, or correlations between the patterns in the data and organize pattern into categories from these correlations.

The perceptron is the simplest form of neural network used for the classification of pattern which are linearly separable as shown in "Fig. 4". Basically it consists of a single neuron with adjustable synaptic weights and bias. Perceptron learning algorithm adjusts the free parameters of the neural network during training phase. The classes have to be linearly separable for the perceptron to classify properly. In the "Fig. 3" the synaptic weights w_1, w_2, ...w_n of the perceptron, correspondingly, the inputs applied to the perceptron are denoted by x_1, x_2, ...x_n with the externally applied bias denoted by 'b' with the output 'v' is represented as follows

$$v = \sum_{i=1}^{m} w_i x_i + b \qquad (1)$$

The goal of the perceptron is to correctly classify the set of externally applied stimulus x_1, x_2, ...x_n into one of two classes C1 and C2. The decision rule for the classification is to assign the point represented by the inputs x_1, x_2, ...x_n to class C1 if the perceptron output is +1 and to class C2 if it is -1. The bias $b (n)$ is treated as a synaptic weight with a weight +1, thus the input vector is defined as

$$X (n) = [+1, x_1, x_2, ...x_n]^T$$

Here 'n' denotes the iteration step. The associated weight vector is given by

$$W (n) = [b(n), w_1, w_2, ...w_n]^T$$

Multilayer perceptrons (MLP), with input layer, hidden layer and output layer is used for solving the classification
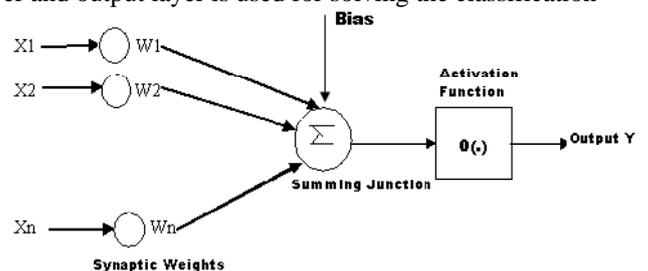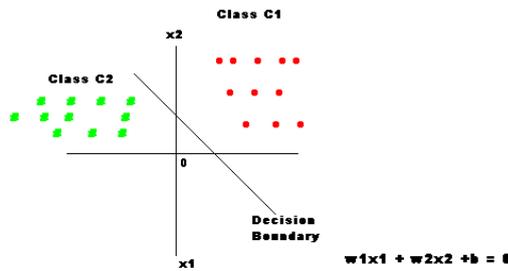

Fig. 3. Perceptron

Fig. 4. Linearly separable classes

Problem proposed in this paper by training the network in a supervised manner with a highly popular error back propagation algorithm. The task of feature extraction is performed by hidden layer of the network. Given a set of example pairs (x, y), where x €X and y €Y a function

$$f: X \rightarrow Y$$

With a minimized cost function $C = E (f(x)-y)$ for data pairs (x, y).

In a multilayer perceptron, the error signal at the output of neuron 'j' at iteration n is defined by

$$e_j(n) = d_j(n) - y_j(n) \qquad (2)$$

where neuron j is an output node, $d_j(n)$ is the desired output and y_j (n) is the actual output and $e_j(n)$ is the error signal at the output. The instantaneous error energy for neuron j is

$$\mathcal{E}(n) = 1/2 \sum_{j=c}^{n} e_j^2(\text{n}) \qquad (3)$$

Here c includes all the neuron in the output. Let n denote the total number of patterns contained in the training set. The average squared energy is obtained by summing $\mathcal{E}(n)$ over all n and then normalizing with respect to the set size N.

$$\mathcal{E}(n) = 1/N \sum_{n=1}^{N} \mathcal{E}(n) \qquad (4)$$

IV. PROPOSED ALGORITHM

In this paper using the algorithm proposed in [6], input color image is first skin segmented using HSI and YCbCr color spaces. For a detailed survey of color spaces refer [11]. As all the skin segmented regions are not face regions, each segmented region has to be checked whether the segmented region contains face or not the following algorithm is proposed. The proposed algorithm precisely locates where exactly the actual face lies in each skin segmented region as the entire segmented region does not represent the face. This algorithm is implemented in two phases namely Training phase and Testing phase.

A. Training Phase

For feature extraction, images with only dominant facial features eyes, nose and mouth with variations in pose and expressions which fit in to a window of size 88 X 80 are considered.

Step-1: Wavelet decompose each histogram equalized face image containing prominent facial features with slight variation in pose and expression up to second level, retain only second level approximation image and discard the

horizontal, vertical and diagonal details .

Step-2: Apply Canny and Sobel edge detection algorithm separately on the second level approximation images of each training image considered for feature extraction.

Step-3: Combine the edges of the approximation image obtained in Step-2, into a single edge image using pixel by pixel image multiplication. On this edge image perform morphological operations such as erosion and dilation to extract the prominent facial edge features.

Step-4: The extracted prominent facial edges called wavelet edges are stacked into a data edge matrix as column vectors.

Step-5: Similarly windows not containing human faces are also wavelet decomposed and apply Step-2 to Step-3 to obtain non-facial edges. Append these edges to the data edge matrix of Step-4. This serves as the input data matrix for training the neural net work

Step-6: Create an output column vector having the number of rows equal to the number of columns of input data edge matrix created in the previous steps. The output column vector values are initialized to +1 or 0 indicating face or non-face.

Step-7: Initialize the neural network parameters and train the network to obtain appropriate weight sequence w(n) using back propagation algorithm

Testing Phase

Step-1: Skin segment the input color image using the algorithm proposed in [12]. For skin segmented regions with size larger than the window size considered, use sliding window technique proposed in [12].

Step-2: Histogram equalize each window image followed by wavelet decomposition up to second level. Retain only second level approximation image.

Step-3: Extract the edges of the second level approximation image using Step-2 and Step-3 of training phase.

Step-4: Test these edges with the previously trained neural net

V. RESULTS

The In the experiments 100 face images and 75 non-face images of size 88 X 80 were used for extracting second level approximation feature set edges. Group images containing faces with variation in pose, having moustache, structural components and slight variation in expressions and non-faces were used for testing. Rowley *et al*. [13] have used around thousand images for training, where as in the proposed method, as facial wavelet-edge features are already extracted, the small training set is sufficient. This approach also detects faces with variation in pose and structural components. In [14] approximation coefficients were used without extracting edges. The number of false acceptances was higher in skin tone regions and there were false rejections in skin tone images with complex background. In [12] the number of false acceptances and false rejections were reduced considerably when the edge strength of the second level approximation was used as the feature set.

In [6], there are few false rejections and the number of false acceptance is minimized as the window contents are checked for the presence of face in two different channels (R

and G) for matching edge strength. False acceptances are noticed only when approximation images produce similar edges at identical positions. In the proposed method, multi layer perceptrons are used for classification, instead of city block distance measure for classification. The experiment was conducted on 3000 window contents containing faces and non-faces. False positives and false rejections found to be less than 1% and the result is highly encouraging. Window contents of "Fig. 5a" which was rejected as a face due large variation in facial expression is correctly classified as a face by the proposed algorithm, similarly "Fig. 5b" which was falsely classified as a face by [15] due to the edges produced at identical positions, is correctly classified in this method. Classification using neural network on both R and G channels is expected to further improve the detection. Comparative results are tabulated in Table I.
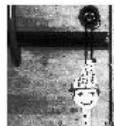

Fig. 5a   Fig. 5b

TABLE 1: RESULTS USING DIFFERENT METHODS

Method	Classification Method	Windows Tested	False Positive	False Rejection
Proposed Methodology	Neural Network	3000	5	0
Approximation image with R and G channel Edges	Cityblock	2800	18	8
Approximation image with Gray Scale Edges	Cityblock	1700	30	3
Approximation Image without Edge Extracion	Bhattacharya Distance	1500	128	20
	Cityblock	1500	44	38

REFERENCES

[1] D. Petrescu and M. Gelgon, "Face detection from complex scenes in colour images."

[2] Y. M. Hsuan, D. J. Kreigman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine intelligence*, vol. 24, pp. 34-58, 2002.

[3] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399-458, 2003.

[4] F. Fritsch, S. Lang, M. Kleinehagenbrock, G. A. Fink, and Sagerer, "Improving Adaptive Skin Colour Segmentation by incorporating results from face detection," in *Proc. IEEE International workshop on Robot and Human Interactive Communication (ROMAN)*, Berlin, Germany, pp. 337-343, 2002.

[5] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," *Second edition Prentice Hall*, India, 2002.

[6] H. C. Vijaylakshmi and S. PatilKulkarni, "Face detection in skin-toned images Using edge detection and feature extraction using R and G channels through wavelet approximations," in. *Proc. International Conference on Machine Vision*, Hong Kong, 2010, pp. 238-241.

[7] S. Krishna. Open CV face detection system. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/19912-open-cv-viola-jones-face-detection-in-matlab.

[8] P. Viola and M. Jones, "Robust real time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2001.

[9] R. Maini and H. Aggarwal, "Study and comparison of various image edge Detection techniques," *International Journal of Image Processing (IJIP)*, vol. 3, no. 1, pp.1-11, 2009.

[10] M. W. Frazier, an Introduction to Wavelets through Linear Algebra, Springer, 1999.

[11] V. Veznevets, V Sazonov, and A Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon-2003*, pp. 85-92.

[12] H. C. Vijaylakshmi and S. PatilKulkarni, "Segmentation algorithm for multiple face detection for color images with skin tone regions," in. *Proc. 2nd International Conference on Signal Acquisition and Processing*, 2010, pp 162-166.

[13] H. A. Rowley, S. Baluja, T. Kanade, "Neural network based face detection", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.20, no. 1, pp. 39-51, 1998.

[14] H. C. Vijaylakshmi and S. PatilKulkarni, "Face localization and detection algorithm for colour images using wavelet approximations," in *Proc. IEEE-ACVIT-09, 2nd International Conference on Advances in Computer Vision and Information Technology*, Dec 2009, pp 305-311.

[15] H. C. VijayLakshmi and S. PatilKulkarni, "Face detection and localization in skin toned color images using wavelet and edge detection techniques," in *Proc. International Conference on Advances in Recent Technologies in Communication and Computing*, 2010, pp. 231-234.

H. C. Vijaylakshmi received her B.Sc degree in physics, chemistry from the University of Mysore, Mysore in 1984, the B.E degree in Computer Science & Engineering from the Universityof Mysore, Mysore in 1988 and M.Tech degrree in Software Engineering from the University of Mysore, Mysore in 1997. She is currently pursuing Doctoral program at University of Mysore, Mysore. She is the author/co-author of 10 publications, in technical journals and conferences. Her area of interest includes Image Processing, Computer Vision, Software Engineering and Database Management Systems.

Sudarshan Patilkulkarni was born on 21st June 1974 at Bagalkot, Karnataka, INDIA. He obtained his Bachelor of Engineering degree (B. E.) from Karnataka University, Dharawad, INDIA, in the year 1996. He earned both Masters in Electrical Engineering (M. E) and Doctor of Philosophy in Electrical Engineering (Ph. D.) from Old Dominion University, Virginia, U.S.A in the year 2000 and 2004 respectively. He is currently working as Scientist, at JSS Research Foundation, Sree Jayachamarajendra College of Engineering, Mysore, INDIA. His areas of research include signal, speech, image processing and control systems.