Face Detection Algorithm for Skintone Images using Robust Feature Extraction in HSV Color Space

H C VijayLakshmi
Dept of Computer Science & Engineering
S J College of Engineering, Mysore,INDIA

S. Patilkulkarni, Ph.D
Associate Professor
S J College of Engineering, Mysore,INDIA

ABSTRACT
A robust feature extraction method in HSV space is proposed for face detection problem in skin toned images using biorthogonal wavelet detail coefficients. It is demonstrated that followed with neural network classifier, proposed method is robust under varying conditions.

General Terms
Feature extraction, threshold, wavelet, color space, skin tone

Keywords
Biorthogonal, HSV color space,

1. INTRODUCTION

1.1 Face Detection
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 face detection problem involves segmentation, feature extraction, and classification of the segmented region as face or non face irrespective of the background and occlusion. Several researchers are at this task with different approaches, so that the machine detects and locates the faces efficiently as we human beings do in any complex scenarios.

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. A large number of factors that govern the problem of face detection [5, 6].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 [5, 6]. In terms of applications, face detection and good localization is an important pre-processing step in online face recognition and surveillance systems. In the recent years lots of progress has been made in detecting faces with slight variation in illumination, pose, expression and background. A number of techniques have been developed by researchers in order to efficiently detect human faces in any given input image or video sequence.

A face detection algorithm with a novel and robust feature extraction method is proposed, which is invariant to illumination, background and expression. Each face image contains wavelet decomposed detail coefficient extracted in hue, saturation, value channels, eye, nostril and mouth blobs are dilated and threshold to obtain facial features. Experiments are conducted using the personal database developed by the authors as well as Bao Face Database available on the internet.

1.2 Wavelets
Wavelets are mathematical functions that initially cut up the data into different frequency components and then study each component with a resolution matched to its scale. Thus wavelets analyze according to scale and self-similarity caused by scales and dilations. Wavelet algorithms process data at different scales or resolutions. The wavelet analysis procedures adopt a wavelet prototype function called an analyzing wavelet or mother wavelet [3]. The Discrete Wavelet Transform (DWT) has become a very versatile signal processing tool over the last decade. In fact, it has been effectively used in signal and image processing applications ever since Mallat [12] proposed the multiresolution representation of signals based on wavelet decomposition. Wavelet transform is a representation of a signal in terms of a set of basis functions, which is obtained by dilation and translation of a basis wavelet. The advantage of DWT over other traditional transformations is that it performs multiresolution analysis of signals with localization both in time and frequency. Orthogonal or biorthogonal wavelet transform has often been used in many image processing applications, because it makes possible multiresolution analysis and does not yield redundant information, refer Mallat [12]. The wavelet consists of two components, the scaling function which describes the low-pass filter for the wavelet transform, and the wavelet function which describes the band-pass filter for the transform. For orthogonal wavelets, the scaling function Φ and mother wavelet Ψ are given by the recursion relations

$$\Phi(x) = \sqrt{2}\sum_k h_k \Phi(2x - k) \qquad (1)$$

$$\Psi(x) = \sqrt{2}\sum_k g_k \Phi(2x - k) \qquad (2)$$

Their scaled translates are denoted by

$$\Phi_k^n\,(x) = 2^{\frac{n}{2}}\Phi(2^n\,x - k) \qquad (3)$$

$$\Psi_k^n\,(x) = 2^{\frac{n}{2}}\Psi(2^n\,x - k) \qquad (4)$$

The translates of scaling function form the basis set for the approximation subspace, where as those of wavelet function form the basis set of detail subspace respectively, at the n-th level resolution. Wavelet transforms can be applied in a number of scientific research areas such as feature extraction, edge and corner detection, partial differential equation solving, transient detection, filter design, electrocardiogram (ECG) analysis, texture analysis, business information analysis and gait analysis. Transforms in image processing are two-dimensional. Discrete wavelet transform is calculated by applying the corresponding one-dimensional transform to the columns first, and then to the rows. When filtering, we have four possibilities

- low-pass filter to rows, followed by low-pass filter to columns (LL coefficients)
- low-pass filter to rows, followed by high-pass filter to columns (HL coefficients)
- high-pass filter to rows, followed by low-pass filter to columns (LH coefficients)
- high-pass filter to rows, followed by high-pass filter to columns (HH coefficients)

In wavelet decomposition, the image is split into an approximation and details images. Approximation image is obtained by low pass filtering and detail images are obtained by high pass filtering. Further decomposing the approximation image (LL1 subband), we will get second level LL2, HL2, LH2 and HH2 coefficients as shown in the Fig. 1a. The high pass or detail component characterizes the images' high frequency information and the low pass or approximation component contains its low frequency information.

Wavelet details contain high frequency components. The more high-frequency components are involved in an edge, the more abrupt the edge transition will be, which corresponds to a clearer edge. Joining multiresolution details in the edge reconstruction procedure will create an effect of enhancing edges. The given input image is wavelet decomposed, the detail coefficients capturing the edge information are considered, the approximation coefficient of the wavelet decomposed image is replaced with the sum of horizontal, vertical and diagonal (HL, LH, HH) detail coefficients of the image. These details contain the pixel intensity change in horizontal, vertical and diagonal directions capturing the important features of the image in all the three directions (Fig. 1b). Images containing different hues but same intensity values have the same pixel intensity value for different in the grayscale image. Color images when converted into grayscale image, may represent different color information with the same pixel intensity, resulting in missing actual image feature. In order to capture the features without losing any vital information, the given color image is converted in to HSV space, so that variation in hue, saturation and value which represent the image features are captured. The image in HSV color space is wavelet decomposed in all the three channels and using the horizontal, vertical and diagonal detail coefficients, the image features in all the three channels which account for variation in hue, saturation as well as intensity are extracted as feature set.

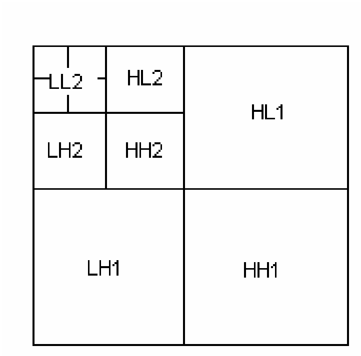

Fig 1a : Multiresolution structure of wavelet decomposition.

Fig 1b: Wavelet decomposition of a face image.

1.3 Preprocessing

Preprocessing step is the stage where the input image is processed to extract the regions and features of interest. The main task of this stage is to avoid processing regions which are not the candidate regions for the algorithm to process and eliminate these regions from further processing. Pre-processing does not increase the image information content. It is useful on a variety of situations where it helps to suppress information that is not relevant to the specific image processing or analysis task (i.e. background subtraction).The aim of preprocessing is to improve image data so that it suppresses undesired distortions and/or it enhances image features that are relevant for further processing.

Segmentation and feature extraction are the two important pre-processing steps that play a vital role in face detection and localisation. Segmentation is a process that partitions an image into regions [7] of uniform texture and intensity. In the problem of face detection, segmentation based on skin plays a major role in identifying the candidate face regions [9]. Though there are different segmentation methods, segmentation based on colour is considered. For a detailed survey of colour spaces refer [11].

Segmentation of the input image is the most important step that contributes to the efficient detection and localization of multiple faces in skin tone colour images Skin segmentation of the input image based on skin pixel cue [8,16,17] combined with edges help to demarcate region boundaries and segment the image components efficiently [8]. Segmentation of an image based on human skin chromaticity using different colour spaces results in identifying even pseudo skin like regions as skin regions. Hence there is a need for further eliminating these pseudo skin regions. The segmentation using combination of colour spaces combined with Canny and Prewitt edge detection for obtaining the region boundaries segment better [9]. To segment the skin regions in color images containing skin tone regions and complex background refer [10]. In this approach, a wavelet texture energy based illumination compensation for segmenting the outdoor bright sunlight skin tone images even with complex background is used. This methodology generates a texture energy image for wavelet approximation image, addresses illumination compensation and results in smooth segmentation. The blurred wavelet texture energy image generated removes small details and bridges small gaps in lines and curves. Random noise typically consists of sharp transitions in pixel intensity level are minimized and thus results in a better object boundary. For each channel approximation image, using a 3X3 image window a texture energy image is generated using the following equation

$$t(x,y) = \frac{1}{h} * \sum_{s=-a}^{a} \sum_{t=-b}^{b} f(x+s, y+t) * f(x+s, y+t)$$

Here f(x, y) is the input image, s and t the size of the window, in our case 3X3 kernel is considered and t(x, y) is the weighted average texture image generated and 'h' is the threshold. Using the approximation texture energy images generated, the image in each channel is reconstructed and the entire three channel images are combined to obtain the illumination compensated RGB image.

2. FEATURE EXTRACTION

Feature extraction is the method of capturing information about the object in compact way, using these features one should be able to retrieve and hence recognize the object under consideration. Features should contain vital information about the object of interest; the feature extraction should be easy to compute and should be robust and compact enough to represent the object under consideration uniquely. The features extracted should well depict the human perception about the object. Image features can refer to global properties of an image i.e. average gray level, shape of intensity histogram etc. and local properties of an image such as edges, textures, important features uniquely representing the object and also shape of contours etc. From object recognition purpose these image features should be local, meaningful, detectable parts of an image. Meaningful features are associated to unique information in the image (e.g., like eyes, eyebrows, mouth and nose in a human face). They should be invariant to some variations in the image formation process (i.e. invariance to viewpoint and illumination for images captured with digital cameras). These features extracted should be detectable, they can be located or detected from images via algorithms and features should robustly capture the salient features of the object. The features extracted should be represented using a feature vector with less complexity. Robust and compact feature set [13, 14, 15] yield good detection rates. In this paper in order to extract facial features without losing any information, color face image is converted in to HSV color space and the variation in hue, saturation and value are also considered.

Edges occur in images due to sudden change in pixel intensity values. Edges are the most important feature of an image. Edge detection is an important task in image processing. It is a main tool in pattern recognition, image segmentation, and scene analysis. An edge detector is basically a high pass filter that can be applied to extract the edge points in an image. Object recognition, image segmentation and image coding applications require robust and smooth edges clearly highlighting the regions of interest. For a detail survey of edges in grayscale images refer [2]. Important features can be extracted from the edges of an image (e.g., corners, lines, curves). These features are used by higher-level computer vision algorithms (e.g., recognition). In gray scale images, the edges correspond to change in illumination, pixel intensity values. This situation is different in case of color images, which gives more detailed edge information. Color plays a very important role in determining object boundaries in color images. In color images, intensity, hue and saturation of a color all play a part in determining object boundaries. Objects having different hues but same intensity values represent the same pixel intensity value in the grayscale image. Color images containing such hues, when converted into grayscale image, may represent different color information in to a single region, resulting in missing actual object feature and boundary. This is very vital for object segmentation, detection and recognition in color images. Edge extraction in color images provide vital clue about the object's hue variation, intensity variation and also the variation in the amount of saturation in hue. In case of grayscale images, the object boundary and also object features are captured when there is a sharp change in the intensity values in a single channel. In case of color images, more information regarding the object can be captured in more than one channel. For a comprehensive analysis refer [1].

3. PROPOSED APPROACH

A holistic approach is used, instead of looking for the presence of facial features such as eye, mouth, nose separately. In the proposed approach, the algorithm uses the following steps to extract the prominent features.

Step1: Extract the facial features by first converting the windows containing face image from RGB color space to HSV color space.

Step2: Obtain the detail coefficients of each of the three channels (H, S, V) after wavelet decomposition using discrete biorthogonal wavelet transform, in particular, 'bior1.3' wavelet filters. The eye and mouth socket regions extracted in H, S, V channel are dilated in all the three channels using morphological operation to enhance the features and make it clearer. Consider only the horizontal, vertical and diagonal details in all the three channels.

Step3: Replace the LL subband image i.e. approximation coefficients with the sum of horizontal, vertical and diagonal details as these detail coefficients capture all the important features of the face. Reconstruct the image with enhanced detail coefficients in each channel.

Step4: The reconstructed image in HSV channel which predominantly contains eyebrows, eyes, nostrils and mouth regions as shown in Figure 2b is converted back to RGB color space and then to grayscale image and an image histogram is generated. Using suitable hard threshold, face images highlighting only the prominent facial features are obtained.

Step5: Using this procedure features are extracted for the facial window images considered for training and stacked as column vectors in a data matrix. Similarly nonface images are also wavelet decomposed in H, S and V channel and using the above procedure, the non facial features extracted are appended to the data matrix. This serves as the input data matrix for training the neural network.

Step6 : Create an output column vector having the number of rows equal to the number of columns of input data matrix created in the previous steps. Initialize the output column vector values to +1 or 0 indicating face or non-face. Initialize the neural network parameters and train the network to obtain appropriate weight sequence *w(n)* using back propagation algorithm. Multilayer perceptrons (MLP), with input layer, hidden layer and output layer is used for classifying the window image as face or nonface. Using this approach, the number of face and non face images required for training is vey less and training is very fast, when compared with other approaches which use neural network for classification.

In testing phase, segment the input image containing multiple faces using the approach proposed by the authors in [10], using a sliding window across the skin segmented regions and convert the window image into HSV color space and apply steps mentioned above to extract the features. Test the extracted features with the neural network trained earlier using features extracted using step1 to step5. Classify the window image as face, if the output is greater than 0.5, otherwise nonface. If the input image contains faces of different size, then the sliding window size is scaled up or down depending on the need and the procedure is repeated to locate faces of different sizes.

The features extracted are as shown in Figure 2. Sample features extracted for windows containing faces with variations in pose and expression used in the experiment are as shown in Figure 3. The images considered for training have upright frontal faces, faces with variation in pose, expression as shown in Figure 4.


Fig 2(a)

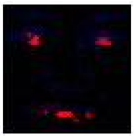
Fig 2(b)


Fig 2(c)


Fig 2(d)


Fig 2(e)


Fig 2(f)

Fig 2(a) and 2(e) Face image; 2(b) Wavelet detail image (horizontal & vertical & diagonal) in HSV color space; 2(c) Converted wavelet detail image (horizontal & vertical &diagonal) in RGB color space; 2(d) and 2(f) Intensity threshold facial features


Fig 3(a)


Fig 3(b)


Fig 3(c)


Fig 3(d)


Fig 3(e)


Fig 3(f)

Fig 3(a) Face with spectacles; 3(b),3(d) and 3(f) Intensity threshold features; 3(c) Face with beard; 3(e) Face with variation in pose.

Fig 4: Sample wavelet detail coefficient Intensity threshold faces used for training.

4. RESULTS

Compared with many other approaches using neural network for classification, proposed approach requires small data set for training and thus consuming less time for training. Around 75 faces and 450 nonface images are used for training. This approach works on low resolution images also. For normal illumination images to extract intensity face features, the threshold range is around 0.04, to extract intensity face features of the low resolution window test images, the threshold range is kept slightly low, say between 0.028 to 0.032.This algorithm can successfully detect faces with variation facial expression, with spectacles, beard and variation in pose and also frontal faces with closed or open eyes. A comparative analysis of detection rates for face detection algorithm using different facial features are tabulated in the Table 1 and Table 2.

5. CONCLUSIONS AND FUTURE WORK

The experiment was conducted on images of personal database developed by the authors as well as, Bao Face Database available on the internet for research purpose. The personal database contains 250 image, around 105 images contain multiple upright frontal faces, around 45 images containing multiple faces with skin tone background and around 100 images with variations in pose. Color images containing multiple faces in skin tone regions and complex background are efficiently segmented by generating wavelet based texture energy image. The proposed method selects only skin regions for with area greater than 800 pixels for classification algorithm. Feature extraction procedure used in this algorithm is computationally efficient and robust. The features extracted using this procedure clearly distinguishes a facial image from a non facial image. We plan to extend this approach to only profile faces as future work. The result is also compared with other approaches [18], [19] proposed by the authors during experimentation and this method is found to have better false rejection and acceptance values.

Table-1 Comparative result of different methods

Features Used	Classification Method	Windows Tested	False Positive	False Rejection	Nature of Dataset Used
Wavelet Approximation coefficients	Bhattacharya Distance	1500	128	20	Frontal Faces Under Normal illumination condition.
	City block	1500	44	38	Frontal Faces Under Normal illumination condition.
Signature of Wavelet Approximation coefficients Image	correlation	2000	50	0	Frontal Faces Under Normal illumination condition.
Gray Scale Edges of Wavelet Approximation coefficients	City block	1700	30	3	Frontal faces with slight variation in pose and expression
R & G channel Edges of Wavelet Approximation coefficients	City block	2800	18	8	Frontal faces with slight variation in pose and expression
Approximation image with Gray Scale Edges	Neural Network	3000	5	0	Frontal faces with variation in pose and expression and spectacles
Proposed Method	Euclidean Distance	3000	100	25	Frontal faces with variation in pose and expression and spectacles.
Proposed Method	Neural network	4000	5	10	Frontal faces with sufficient variation in pose and expressions, faces containing spectacles, Beard, Partly covered faces. Bright sunlight, Normal Illumination Low resolution images.

Table 2 Comparison of approximation image edge features Vs HSV detail image features results

Features Used	Classification method	Windows Tested	False Positive	False Rejection	Nature of Dataset Used
Approximation image with Gray Scale Edges	Neural Network	2000	15	25	Frontal faces with sufficient variation in pose and expressions, faces containing spectacles, Beard, Partly covered faces. Bright sunlight, Normal Illumination Low resolution images.
Proposed Method	Neural network	4000	5	10	Frontal faces with sufficient variation in pose and expressions, faces containing spectacles, Beard, Partly covered faces. Bright sunlight, Normal Illumination Low resolution images.

6. REFERENCES

[1] Shu-Yu Zhu, Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos, Comprehensive analysis of edge detection in color image processing, Optical Engineering, Vol. 38 No. 4, (April 1999), pp 612-625.

[2] Raman Maini, Dr. Himanshu Aggarwal, Study and Comparison of Various Image Edge Detection Techniques, International Journal of Image Processing (IJIP), Vol. 3, Issue 1, pp. 2-12.

[3] L. Feng, C. Y. Suen, Edge extraction of images by reconstruction using wavelet decomposition details at different resolution levels, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 14, No. 6, pp. 779-793.

[4] S. Anila, N. Devarajan, Simple and Fast Face Detection System Based on Edges, International Journal of Universal Computer Sciences, Vol. 1, Issue 2, pp. 54-58.

[5] Y. Ming-Hsuan, D J Kreigman, and N Ahuja "Detecting Faces in Images: a Survey", IEEE Transactions on Pattern Analysis and Machine intelligence, (2002), Vol. 24, pp. 34-58.

[6] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips "Face recognition:A literature survey", ACM Computing Surveys , (2003), pp. 399.458.

[7] R. C. Gonzalez, R. E. Woods Digital Image processing, Second edition Prentice Hall India.

[8] F Fritsch, S Lang, M Kleinehagenbrock, G A Fink and Sagerer "Improving Adaptive Skin Colour Segmentation by incorporating Results from Face Detection", In Proceedings of the IEEE International workshop on Robot and Human Interactive Communication, Berlin, Germany, (September 2002), pp.337-343.

[9] H.C. Vijaylakshmi, S. PatilKulakarni, Segmentation Algorithm for Multiple Face Detection for Color Images with Skin Tone Regions, International Journal of Computer Theory and Engineering, (August, 2010), Vol. 2, No. 4, pp. 552-558.

[10] H.C. Vijaylakshmi, S. PatilKulakarni "Illumination Compensation to Segment True Skin and Non-skin Regions For Skin Tone Images" Springer Computational Intelligence and Information Technology, (2011), Pune Volume 250, 2011, pp 494-499.

[11] V. Veznevets, V Sazonov, nad A Andreeva, A survey on pixel-based skin color detection techniques. In Proceedings of GRAPHICON conference (2003), pp. 85-92.

[12] S. Mallat, "A Theory forMultiresolution Signal Decomposition: TheWavelet Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, (1989), vol. 11, no. 7, pp. 674–693.

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

[14] Rein-Lein Hsu, Mohamed-Abdel Mottaleb, Anil K jain Face Detection in Color Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, Issue 5, pp.696-706.

[15] C. Garcia, G. Zikos, G. Tziritas, Face Detection in Color Images using Wavelet Packet Analysis. In Proceedings of IEEE International Conference on Multimedia Computing and Systems, (1999) Vol.1, pp. 703 – 708.

[16] Mohamed A. Berbar Hamdy M. Kelash, and Amany A. Kandeel, "Faces and facial features detection in color images," In Proceedings. Geometric Modeling and Imaging— New Trends (GMAI'06), pp 209 – 214.

[17] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai, Skin segmentation using colour and edge information. In Proceedings of the International Symposium on Signal Processing and its Applications, Paris, France, (July 2003), pp. 353-356.

[18] H.C. Vijaylakshmi, S. PatilKulakarni, Face Detection in Skin-Toned Images Through Wavelet Edges and Neural Network, International Journal of Computer and Electrical Engineering,Vol. 4, No. 5.

[19] H.C. Vijaylakshmi, S. PatilKulakarni, Face Detection in Skin-toned Images Using EdgeDetection and Feature Extraction Using R and G Channels through Wavelet Approximation, International Journal of Computer Theory and Engineering, Vol. 5, No. 1.