Journal of Engineering and Technology Research Vol. 3(14), pp. 381-387, 22 December, 2011
Available online at http:// www.academicjournals.org/JETR
DOI: 10.5897/JETR11.068
ISSN 2006-9790 ©2011 Academic Journals

Full Length Research Paper

Facial feature units' localization using horizontal information of most significant bit planes

Asif Khan[1]*, Khalilullah[1], Ihtesham-Ul-Islam[1] and Mohammad A. U. Khan[2]

[1]FAST National University of Computer and Emerging Sciences, Peshawar, Pakistan.
[2]Effat University, Jeddah, Saudi Arabia.

Accepted 8 November, 2011

We present here an approach to find the exact position of some feature units related to human face images. We use the horizontal information in most significant bit planes of images to accomplish the task. Finding location of facial feature units is of importance as most human face recognition approaches take it as initial point. The prominent feature units in a face are eyes, nostrils and lips which are usually oriented in horizontal direction and visually significant in face image. The majority of the visually significant data in image can be extracted using higher order bits of that image. Our four step method consists of bit planes processing, separating horizontal information using wavelet transform (WT), binary thresholding and appropriate combination of Dilation and Erosion. The proposed method shows high accuracy in the presence of all real world situations like various gestures, illumination variations, closed eyes, and eyes with glasses.

Key words: Bit planes, feature localization, facial features, face recognition and wavelet transform.

INTRODUCTION

For several decades, extracting facial features and feature units localization are some of the most important research areas in image processing research community as explained by Yang et al. (2002), Lanzarotti et al. (2001), and Yuille et al. (1992). For the success of many applications such as, facial expression analysis, teleconferencing, animation, video surveillance, man machine interfacing, human performance prediction by Jabon et al. (2011), face recognition, lip reading for the deaf and some other tasks, this pre-processing step is essential. In facial expression analysis very high accuracy is required and even very small errors can be interpreted differently and can lead to wrong facial expression recognition. To locate the exact position of facial feature units is imperative in three dimensional modelling, texture mapping and the subsequent animation. For a complete facial image processing system to work, it should be able to do some tasks in a sequence. Detection of faces in a given image, extraction of facial features, recognizing people and describing facial expressions are some of those tasks. In this paper, our focus is on finding the location of various facial features which is important for the remaining two components of the system. Factors that add up to the computational cost for facial feature unit localization are locating the face region, complexity of the evaluation of cost functions, and the searching for the feature points.

*Corresponding author. E-mail: asif.khan@nu.edu.pk. Tel: +92(091)111128128. Ext. 104. Fax: +92(091)5822320.

However face region extraction is no more a problem as there are many robust face detection algorithms available such as Viola and Jones (2001).

In surveys, there are four classes of facial features extraction starting with the use of geometrical information of facial features as first class, proposed by Xhang and Lender (2000) and Rizon et al. (2000). Spors and Rabenstein (2001) and Perez et al. (2001) proposed the spatio-temporal information-based techniques to extract the features from the image sequence. For the third class, the color information of each facial feature is used by Hsu et al. (2002) and Xin et al. (1998). In fourth class, important parameter for detection is intensity of the image as stated by Mariani (1999) and Chandrasekaran et al. (1997). Most facial features extraction methods are not robust and are sensitive to noise, uneven illumination, variations in orientation, colour space used and gestures. The existing algorithms for facial features extraction such as stated by Lavagetto and Curinga (1994), Desilve et al. (1995), Huang et al. (1993), Hess and Martinez (2004), Yuille et al. (1992), Zuo et al. (2004) and Majumder et al. (2011) are unable to solve the problem completely in an efficient way. Due to its high complexity, feature units localization is still under research by a lot of researchers in this area. A related work has been performed by Jun et al. (2003) but that cannot take all the real world situations.

Our objective in the current research and this paper is to find the estimated position of eyes, nose and lips in a facial image taken in any of the real world situation. The idea is to utilize the horizontal information in most significant bit planes after decomposing the face image into bit planes. The decomposition of grey image in to bit planes shows that majority of the visually significant data are available in higher order bits. In a typical face image the visually significant data are eyes, nose and mouth. It is also evident that these feature units are also horizontal information in face image. Our method consists of extracting information contained in upper bit planes of the image, concentrating on horizontal information, binary thresholding and appropriate combination of Dilation and Erosion. This method detects the eyes, nose and mouth in a digital photograph of human face. The main advantage of our method is to remove trade-off between complexity and tackling all real world situations. In section 2, we provide explanation of our proposed method. Section 3 provides simulation results and discussion about our work and section 4 provides conclusion.

MATERIALS AND METHODS

Our proposed method revolves around the idea that facial feature units are significant and horizontal data in a face image. As described in introduction that facial feature units are visually significant data in face image and we can extract this data using Most Significant Bit (MSB) planes of the image. Second part of the idea is that these feature units are horizontally oriented. We can use Wavelet Transform (WT) for the extraction of this horizontal information. Our proposed method can be divided in to four steps. First step consists of extracting MSB planes of the face image. We combine some MSB planes and use the resultant image for facial features extraction. The image obtained in the first step is processed using Wavelet Transform (WT), and this is the second step. In this step we retain only those WT coefficients which represent horizontal information and discard other coefficients to focus on mouth, nose and eyes in the image. Third step consists of thresholding to find the estimated regions of stated features in the image. Fourth step consists of combination of dilation and erosion as post-processing for increasing accuracy of the estimation. After the four steps, we draw bounding boxes around the eyes, nostrils and mouth regions in the original image.

Image or 2D signal is represented in digital form as a matrix with pixels having one of 2^N possible values, defined by N bits. Each pixel can be decomposed into N binary values from MSB to LSB and thus the whole image can be treated in this way, thus forming N bit planes, explained by Pratt (1978). These bit planes are loosely related with N subband images, where subband image with lowest spatial frequencies correspond MSB plane, and LSB plane corresponds to that of highest spatial frequencies. Depending on the values of N bits, each pixel will have corresponding value in each bit plane. Original image, $x(i,j)$, can be reconstructed from its bit planes $b_i(i,j)$ as:

$$x(i,j)=\sum_{k=j}^{N} 2^{N-K}\cdot b_k(i,j). \quad (1)$$

The number of planes that can be obtained with N = 8 bits per pixel are 8, each of them having the size of the original image and with pixels having two values: either 0 or 1. In our experiments the size of the original image is 256 x 256. For demonstration purposes the decomposition of image Lena in to bit planes is shown in Figure 1. The bit planes are arranged in the ascending order from MSB $b_1(i,j)$ to LSB $b_8(i,j)$. The fact is clear that MSB image $b_1(i,j)$ has large areas with uniform values (0 or 1). It is also evident that all the main features of the image are present in four MSB images and the lesser significant bit planes tend to have no characteristics and do not contribute much information.

Image analysis and processing through bit planes in such a way that extract bit planes with main horizontal features in face image is the basic idea in this paper. In a typical face image the features are eyes, nose and mouth and they are also horizontal. The other features like hair, ears etc. are not very useful for most of applications and we do not detect them in this work. We can use one or combination of more than one MSB planes for facial features extraction. Extensive experiments have shown that combining two

Figure 1. Lenna image (top centre) and its eight bit planes, from MSB plane (upper left) to LSB plane (bottom right).

Figure 2. A typical face image with glasses from Yale atabase.

Figure 3. Combination of two MSB planes for the face image in Figure 2.

MSB planes concentrate only on the visually significant features in face image. Thus we combine the two MSB images and process it further for accurately identifying eyes, nose and mouth regions. A typical face image from Yale database is given in Figure 2 and image formed by combining its two MSB images is shown in Figure 3. Removal of glasses without disturbing eyes is one of the major achievements of this step.

Extracting horizontal information

Combining two MSB planes discards all the irrelevant components and intensities from the face image and retains significant feature

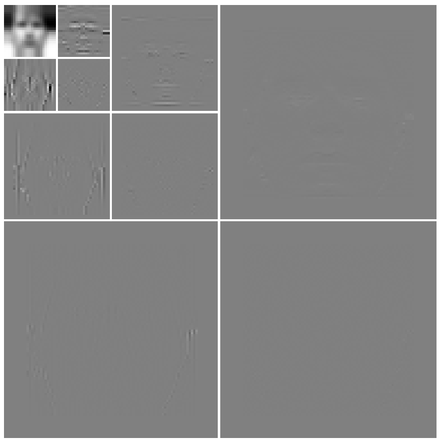

Figure 4. A three scale wavelet transform (WT) of face image.

Figure 5. Horizontal information of face image in Figure 3, after WT processing.

units. Now its turn for the second part of our idea, which states, that lips, nostril and eyes are all horizontal features in a face. Now, we need to eliminate all features except horizontal features from the image obtained in first step. We can do this using Fourier Transform (FT), Wavelet Transform (WT) or Gabor Transform (GT). We have more control over WT coefficients; therefore, we use WT in this paper for extracting horizontal information. Wavelet is, basically, a function $\psi \in L^2(R)$ with zero average:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{2}$$

We can then define the continuous wavelet transform (CWT) of a signal x(t) as:

$$CWT_{\psi}x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^{*}\left(\frac{t-b}{a}\right)dt \tag{3}$$

Where ψ(t) is called the mother wavelet with a and b are called the scaling and translation parameters, respectively. Discrete Wavelet

Figure 6. Result of binary thresholding.

Transform (DWT) is given by

$$DWT_{\psi}x(m,n) = \int_{-\infty}^{\infty} x(t)\psi_{m,n}^{*}(t)\,dt \tag{4}$$

Where

$$\psi_{m,n}(t) = a_0^{-m/2}\psi\left(a_0^{-m}t - nb_0\right) \tag{5}$$

The transforms in one-dimensional form are easily extended to transforms in two-dimensional forms like images. We require a two-dimensional scaling function, φ(x, y), and three two-dimensional wavelets, $\psi^H(x, y)$, $\psi^V(x, y)$, $\psi^D(x, y)$ in two-dimensional case. The wavelets have the ability to measure variations along columns, rows, and diagonals respectively. A three-scale WT can be seen in Figure 4. We retain horizontal information in Figure 3 and discard other coefficients of WT. Then we recover the intensity image which contains only the horizontal information, as can be seen in Figure 5.

Thresholding

By using thresholding we convert image to black and white where

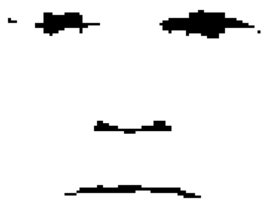

Figure 7. Result of dilation on thresholded image.

Figure 8. Result of erosion on dilated image.

black components represents feature units of interest i.e. eyes, nose and mouth. We also crop some number of upper and lower rows and left and right columns form this binary image because we know that eyes and mouth remain in the center part of the face. The result of thresholding can be seen in Figure. 6. In this step some small irrelevant spots appear near eyes, nose and mouth. We can remove these small areas by appropriate use of dilation and erosion.

Dilation and erosion

In a typical face image only the edges of eyes, nostrils and lips contain the horizontal information. But due to illumination variationsand various gestures there are some false horizontal information on the face. This false horizontal information is small as compared to eyes, nose and mouth. To remove this false information we use combination of two morphological operations, first dilation and then erosion in the next step.

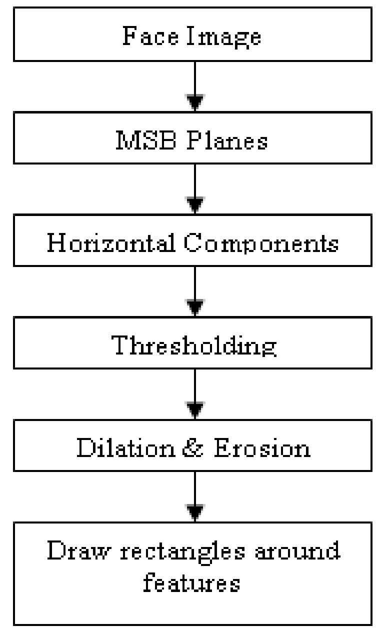

Figure 9. Sequential block diagram of the proposed method.

If A and B are two sets in Z^2, the dilation of A by B, denoted as $A \oplus B$, is defined as:

$$A \oplus B = \left\{ z \mid \left(\hat{B} \right)_z \cap A \neq \emptyset \right\} \qquad (6)$$

Set B is usually called as the structuring element for dilation.

The erosion of A by B, denoted as $A \otimes B$, is defined as:

$$A \otimes B = \{ z \mid (B)_z \subseteq A \} \qquad (7)$$

Figure 7 show the result of dilation where the black areas exactly represent the eyes, nostrils and mouth regions in face image of Figure 2. Figure 8 shows the effect of erosion on Figure 7. In our case, Erosion connects the separated parts of features. Now there are only four connected components in Figure 8 which are our desired feature units. Figure 9 shows all the steps of our proposed method.

SIMULATION RESULTS

In all of our experiments, we used the frontal pose face images of Yale face database (1997), Yale face database B

Table 1. Successful results (%).

Database	Single feature unit detected	All feature units detected
Yale	93	91
BioID	91	90
Yale B	90	87

Georghiades et al. (2001), and BioID face database Jesorsky et al. (2001). There are 165 GIF images for 15 persons in Yale Face Database. For 15 persons, there are eleven images per person, one per different facial expression or configuration. These configurations are center-light, w/glasses, happy, left-light, w/no glasses, sad, normal, right-light, sleepy, surprised, and wink. The Yale face database B consists of 5850 images for 10 subjects, each taken under 576 viewing conditions (nine poses and sixty-four illumination conditions). We used only frontal pose face images for our experiments. The BioID dataset contains 1521 grey level images for 23 different persons. Each one shows the frontal view of a face with resolution of 384x286 pixels. These databases are considered the difficult one because they contain various illumination conditions. In the experiments we used WT up to fifth level and at each level we retained only the horizontal information and discarded the other coefficients. The size of structuring element used is 3x5 for horizontal information. This size may vary for different images. The simplest value for thresholding is the mid value of minimum and maximum intensities in the image of previous step.

We have also done our experiments on faces with glasses and found successful results of our proposed method. The method gives accurate results even in the presence of various gestures of eyes and mouth and also for closed eyes. Table 1 shows the outcome of our proposed method on the above mentioned three databases. Figure 10 shows some of the successful results.

Figure 10. Some successful results of features extraction in face images, before processing (Left) and after processing (Right).

DISCUSSION

The purpose in this study is to find the location of eyes, nose and lips in a facial image taken in any of the real world situation. The idea is to utilize the horizontal information in most significant bit planes after decomposing the face image into bit planes. Our method consists of extracting information contained in upper bit planes of the image, concentrating on horizontal information. Conceptually this work has some advantages and disadvantages. The main advantage of our method is to remove trade-off between complexity and tackling most of the real world situations. The disadvantage is that its accuracy is greatly affected by a variation in face pose. In the future, we hope that, this work can be used successfully to develop more efficient and robust techniques.

REFERENCES

Chandrasekaran V, Liu ZQ (1997). Facial features detection using compact vector field canonical templates. IEEE Int. Conf. on Systems, Man and Cybernetics, 3: 2022-2027.

Desilve LC, Aizawa K, Hatori M (1995). Detection and tracking of facial features in SPIE Symp. Visual Communications and Image Processing, Taipei, Taiwan, R.O.C, pp. 1161-1172.

Georghiades AS, Belhumeur PN, Kriegman DJ (2001). From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," IEEE Trans. PAMI, 23(6): 643–660.

Hess M, Martinez G (2004). Facial features extraction based on the smallest univalue segment assiminating neuclus (susan) algorithm. Picture Coding Symposium, California, USA.

Hsu RL, Abdel-Mottaleb M, Jain AK (2002). Face detection in color images. IEEE Trans. PAMI, 24(5): 696-706.

Huang HC, Ouhyoung M, Wu JL (1993). Automatic feature point extraction on a human face in model-based image coding. Optical. Engineering, 32: 1571-1580.

Jabon ME, Sun JA, Bailenson JN (2011). Automatically Analyzing Facial-Feature Movements to Identify Human Errors. IEEE Intelligent Systems, 26: 54 - 63.

Jesorsky O, Kirchberg K, Frischholz R (2001). Robust Face Detection Using the Hausdorff Distance," In J. Bigun and F. Smeraldi, editors, AVBPA 2001, 90-95. Springer. Available: http://bioid.com.

Lanzarotti R, Campadelli P, Borghese NA (2001). Automatic Features Detection for Overlapping Face Images on Their 3D Range Model. Proceedings of the IEEE 11th Int. Conf. on Image Analysis and Proceeding (ICIAP2001), Palermo, Italy, 26-28, 316-321.

Jun M, Hong L, Wen G, Hongming Z, Gang D, Xilin C (2003). A System for Human Face and Facial Feature Location. Int. J. Image Graphics. 3(3): 461-479.

Lavagetto F, Curinga S (1994). Object-Oriented scene modeling for interpersonal video communications at very low bit-rate. Signal Processing, Image Communication, 6: 379-395.

Majumder A, Behera L, Subramanian VK (2011). Automatic and Robust Detection of Facial Features in Frontal Face Images. UK Sim 13th International Conference on Computer Modelling and Simulation, UK.

Mariani R (1999). Subpixellic eyes detection. IEEE Int. Conf. on Image Anal. Proc., pp. 496-501.

Perez CA, Palma A, Holzmann CA, Pena C (2001). Face and eye tracking algorithm based on digital image processing. IEEE Int. Conf. on Systems, Man and Cybenetics, 2: 1178-1183.

Pratt W (1978). Digital Image Processing, J.W. & Sons, New York.

Rizon M, Kawaguchi T (2000). Automatic eye detection using intensity and edge information. TENCON 2000. Proceeedings, 2: 415-420.

Spors S, Rabenstein R (2001). A real-time face tracker for colour video. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 3: 1493-1496.

Viola P, Jones M (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. Computer Vision and Pattern Recognition Conf., pp. 511-518.

Xhang L, Lender P (2000) Knowledge-based eye detection for human face recognition. 4th IEEE Int. Conf. on Knowledge-Based Intelligent Eng. Sys. Allied Tech., 1: 117-120.

Xin Z, Yanjun X, Limin D (1998). Locating facial features with color information. IEEE Int. Conf. on Sig. Processing, 2: 889-892.

Yale face database (1997). http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

Yang MH, Kreigman DJ, Ahuja N (2002). Detecting Faces in Images: A Survey. IEEE Transactions on PAMI, 24(1): 34-58.

Yuille A, Hallinan P, Cohen D (1992). Feature extraction from faces using deformable templates. Int. J. Comp. Vision, 8(2): 99-111.

Zuo F, de With PHN (2004). Real-time facial feature extraction by cascaded parameter prediction and image optimization. Proc. ICIAR, pp. 651-659.