Face Region Detection Using DCT and Homomorphic Filter

Kyunghwan Baek[1], Yeongsoo Chang[1], Daeuong Kim[1], Youngtak Kim[1], Boohyoung Lee[2], Hwanik Chung[3], Youngjoon Han[1], Hernsoo Hahn[1]
[1]School of Electrical Engineering, Soongsil University
[2]School of Computer Science & Engineering, Kongju National University
[3]Dept. of Internet Information, Kyungbok College
[1]1 Sangdo-Dong, Dongjak-Ku, Seoul, 156-743, Korea
[2]275 Budae-Dong, Cheonan City, Chungnam, 330-717 Korea
[3]131 Sinbuk-myeon, Shinpyoung-ri, Pocheon-si, Gyeonggi-do, 487-910 Korea
[3]http://internet.kyungbok.ac.kr/hichung/home

Abstract: - This paper proposes a new approach for the detection of face regions where the facial features such as eyes, nose, and mouth may exist with very high probability. It uses the characteristic of DCT(discrete cosine transformation) that concentrates the energy of an image into lower frequency coefficients. Since the facial features are pertained to relatively high frequency in a face image, the inverse DCT after removing the DCT's coefficients corresponding to the lower frequencies generates the image where the facial feature regions are emphasized. For an efficient application of the proposed method, homomorphic filter is used with the global contrast factor. The DCT coefficients are eliminated so that the global contrast factor can be maximized. Once the inverse DCTed image is obtained, the face region is found by analyzing the topology of facial features. The proposed algorithm has been tested with various images collected randomly. The experimental results have shown a superior performance even when an image has a complex background.

Key-Words: - Facial features and face detection, DCT, Homomorphic Filter.

1 Introduction

Recently, biometric technologies have been widely used for implementing security and surveillance systems. Such biometric technologies can be classified into two types, controlled and uncontrolled, depending on the data acquisition methods of input devices. Controlled type requires a strong support from the users to acquire the data and uncontrolled type does not. Controlled type includes the systems recognizing fingerprint, iris, venous blood, while uncontrolled type includes the systems identifying face and speech.[1,2] The accuracy and reliability of controlled type recognition technologies have been approved already and thus most security systems have adopted this type of techniques. Contrary to controlled type ones, the uncontrolled type technologies do not require the users to contact the input devices and thus may collect input data with much less help from the users. The characteristic of collecting data without contacting the users makes these technologies very attractive for the surveillance systems, although they have relatively lower reliability and accuracy compared to the contact type systems, due to inevitable noise in input data.

This paper deals with a face region detection technique which can be classified as an uncontrolled biometric scheme. Many researchers have proposed various approaches for this problem, which should be accomplished prior to face recognition process. Recently, Yang[3] has classified these approaches into four categories: knowledge-based, feature invariant, template matching and appearance-based approaches. Among these approaches, feature invariant ones have been used in popular, where the face region is determined by searching the facial features[4] such as eyes, eyebrows, nose and mouth, as well as skin color[5] and textures[6] which exist in any face regardless of sex, age, and race. One difficulty of using these approaches is that detection of the facial features is significantly influenced by lighting condition and image size. To solve this difficulty, many researchers have introduced various methods. Basri[5] and Ramamoorthi[6] have respectively represented convex Lambertian objects using 9-D linear subspace and shown that the images of convex Lambertian objects acquired under various lighting conditions can be approximated by the harmonic images spanned by 9 eigen-images. These methods have shown much improved performance in recognizing faces under irregular lighting conditions. However, they suffered from increased computations due to complexity of the algorithm to generate harmonic images and from the difficulty in applying to small images since the facial features cannot be extracted easily when image size is small. Viola and Jones[7] presented a new visual object detection scheme which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. Although this method increases the speed of the detector by focusing attention on promising regions of the image, many operations per pixel to compute 'Integral Image' are needed and many classifier stages are required for the high detection rate .

To solve such problems of performance deterioration depending on lighting conditions and face colors, this paper proposes a new method of detecting face regions in a gray level image using the characteristic of DCT(discrete cosine transformation) which concentrates the energy of the image on the lower frequency coefficients. Since the facial features occupy the higher frequency regions in a face image, the facial feature regions will remain by taking the inverse DCT with eliminating the lower frequency coefficients of DCT. For maximize the effect of this operation, the homomorphic filter[8] is used with the global contrast factor[9], as shown in Fig. 1 which shows the flow of the proposed algorithm. The experimental results have shown that the proposed algorithm works robust on the variation of lighting condition and face size and background.

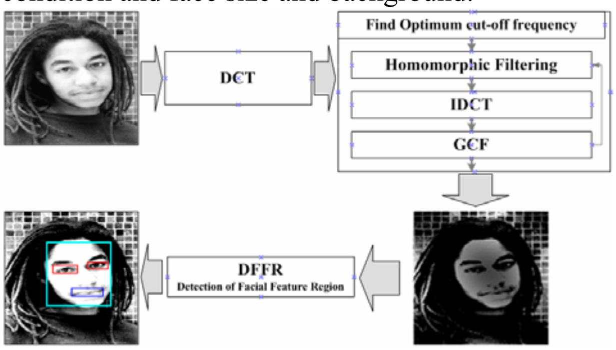


Fig.1. The flow of the proposed algorithm.

2 Detection of Face Regions using the Characteristic of DCT

2.1 DCT(Discrete Cosine Transformation)

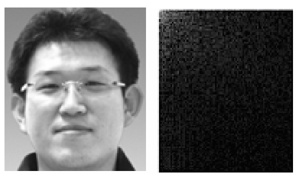
DCT is one of the transforms, such as DFT(Discrete Fourier Transform), which map the signal in time domain into frequency domain. Differently from other transforms whose eigen-functions are complex ones including imaginary terms, DCT uses the real functions as the eigen-functions as given in Eq. (1) and concentrates the energy of an image into the coefficients of lower frequency components.

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2N}\right) \quad \textbf{(1)}$$

Where $f(x,y)$ is the input image, and its coefficients $\alpha(u)$ and $\alpha(v)$ are given as follow:

$$\alpha(u),\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & for\ u,v=0 \\ \sqrt{\frac{2}{N}} & for\ u,v\neq 0 \end{cases} \quad \textbf{(2)}$$

The coefficient of the lowest order in DCT, $F(0,0) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)$, represents the DC component which is the average of the image. As the order increases, the coefficient represents the higher frequency component. Fig. 2(b) shows the result of DCT applied to the face image given in Fig. 2(a). It can be found easily from Fig. 2(b) that only the coefficients of lower order have the large values and thus the image can be approximated with only the coefficients of lower order. This is why DCT is popularly used for image compression.



(a) (b)

Fig.2. A face image and its DCT image

2.2 Inverse DCT with removing the coefficients of lower order

When an input image is small or lighting condition is irregular, any edge operator cannot properly detect the boundaries of the facial features since they become blurred in such cases and false edges can be generated. To overcome these problems, this paper proposes a new method of taking DCT of the input image and taking inverse DCT with eliminating the coefficients of lower orders corresponding to the lower frequencies. In the result image, the facial feature regions are emphasized and the boundaries of the facial features can be detected easily.

In Eq. (3) describing the inverse DCT, $F(u,v)_{rem}$ is the set of the DCT coefficients where the coefficients corresponding to the DC and lower frequency components are eliminated.

$$f(x,y)_{rem} = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)F(u,v)_{rem} \cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2N}\right) \quad \textbf{(3)}$$

Then the histogram of inverse DCTed image becomes to have a bimodal shape as shown in Fig. 3 where the higher part is generated by the facial feature regions and the lower part is formed by the background or other features.

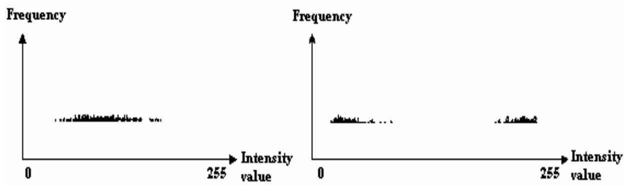


Fig.3. Comparison of the histograms of the inverse DCTed images without (a) and with (b) removing the coefficients of lower frequency components.

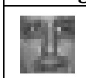
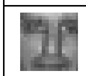
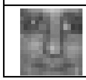
Original face image	IDCTed Image with removing DCT coefficients			
	~ F(1,1)	~ F(2,2)	~ F(3,3)	~ F(4,4)
				
				
				

Fig.4. The results of IDCT with removing the lower frequency DCT coefficients.

Fig. 4 shows three images of 16x16 pixels and their IDCTed images obtained by applying Eq. (1) and (3) in sequence. In the figure, the first column shows the case where the DCT coefficients from F(0,0) to F(1,1) are eliminated and the inverse DCT is followed. The second one does the case where the DCT coefficients from F(0,0) to F(2,2) are eliminated, the third one and the fourth one respectively show the cases where the DCT coefficients from F(0,0) to F(3,3) and F(4,4)are eliminated. It can be found in the figures that the relatively brighter regions include the facial features such as eyes, nose, and mouth. Since the histogram of the result image has a complete bimodal shape as given in Fig. 3, the feature regions and can be easily segmented.

3 Removing DCT Coefficients based on Homomorphic Filtering

In order to determine up to which order of DCT coefficients should be removed, two factors has been considered. The first one is that the facial feature regions can be clearly segmented from other area, and the features in the facial feature regions may appear clearly. For this purpose, the homomorphic filter is used.

Homomorphic filter has been used with Discrete Fourier Transform to emphasize the contrast of an input image. The homomorphic filter, given in Eq. (5) is applied in the frequency domain, as shown in Fig. 5. It is assumed that an input image is the multiplication of illumination and reflectance components, and the contrast in the image can be emphasized by adjusting D_o in Eq. (5).

$$\ln[f(x,y)] = \ln[i(x,y)] + \ln[r(x,y)] \quad \textbf{(4)}$$

$$H(u,v) = (\gamma_H - \gamma_L)[1 - e^{-c(D^2(u,v)/D_0^2)}] + \gamma_L \quad \textbf{(5)}$$

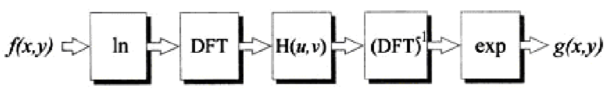


Figure 5. Homomorphic filtering with DFT.

This filter is used in this paper to determine which DCT coefficients should be eliminated to obtain the image with the best contrasted image. As shown in Fig. 1, a global contrast factor(GCF) is used as the measure of contrast. Once the DCT image is obtained, the homomorphic filter is applied with $\boldsymbol{D_o}$ of 0.1 and the GCF of the inverse DCTed image is calculated. This process is repeated with changing $\boldsymbol{D_o}$ up to 4.5. The $\boldsymbol{D_1}$ whose GCF is the largest is selected and used for determining the DCT coefficients to eliminate. The DCT coefficients are eliminated by setting the DCT (u, **v**) as given in Eq. (7).

$$D_1 = \sqrt{u^2 + v^2} \quad \textbf{(6)}$$

$$DCT(u,v) = \begin{cases} 0 & if \quad D_1 < D_0 \\ 1 & \quad\;\; D_1 > D_0 \end{cases} \quad \textbf{(7)}$$

The global contrast factor is the weighted average of the contrast factors with various image resolutions. For an image whose resolution is γ , Contrast is measured based on luminance. Luminance l of a pixel is determined by gamma correction γ of a pixel value as given in Eq. (8), and its perceptual luminance ***L*** is calculated by Eq. (9)

$$l = (\frac{k}{255})\gamma \quad \textbf{(8)}$$

$$L = 100 * \sqrt{l} = 100 * \sqrt{(\frac{k}{255})\gamma} \quad \textbf{(9)}$$

Then the local contrast(***lc$_i$***) in a region whose size is w centering the pixel i is determined by the following Eq. (10), and the contrast measure of the image whose resolution is r is the average of the local contrasts as given in Eq. (11).

$$lc_i = \frac{|L_i - L_{i-1}| + |L_i - L_{i+1}| + |L_i - L_{i-w}| + |L_i - L_{i+w}|}{4} \quad \textbf{(10)}$$

$$C_i = \frac{1}{w*h} * \sum_{i=1}^{w*h} lc_i \quad \textbf{(11)}$$

To finally determine the global contrast factor, the contrast measures are calculated with changing the resolution of the image. If the number of resolutions is N, then the global contrast factor of an image is determined by the following Eq. (12), where the weight is determined by Eq. (13).

$$GCF = \sum_{i=1}^{N} w_i * C_i \quad \textbf{(12)}$$

$$w_i = (-0.406385 * \frac{r_i}{9} + 0.334573) * \frac{r_i}{9} + 0.0877526 \quad \textbf{(13)}$$

In Eq. (13). ***i*** is the index of resolution and r_i is the resolution corresponding to the index.

4 Detection of Facial Feature Region by Testing Faceness

Once the IDCTed image is obtained, it should be tested whether the facial feature regions actually include the features. Since it is impossible to make a crisp decision on the individual facial features, here the topological relations among the facial features are tested.

In the first step, all possible features are selected inside the facial feature regions and then it is tested whether they satisfy the statistically generated topological template of a face, as shown in Fig. 6. In the second step, the possible combinations of these features are evaluated to select the sets of features satisfying the topological rules of a face. Among the features, those are selected if they form the shape given in Fig. 6 and satisfying the condition given in Eq. (13).

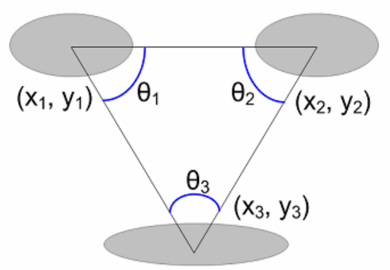


Fig. 6. Topological template of a face.

$$\theta_{ref}(k) = \begin{cases} 65° & , if \quad k = 1(left\ eye) \\ 65° & , if \quad k = 2(right\ eye) \\ 50° & , if \quad k = 3(mouth) \end{cases} \quad \textbf{(14)}$$

In Eq. (14), sign(·) is the signup function, (x_1, y_1) is the center of the left eye, (x_2, y_2) is the center of the right eye, and (x_3, y_3) is the center of the mouth. To take most faces of common people into account, the angles in the following ranges are considered as a triangle of a face. ***θ_{ref}(k)*** given in Eq. (15), is the reference angle which is derived from the face shapes of common people.

$$\theta_1 = arctang\left(\left|\frac{y_3 - y_1}{x_3 - x_1}\right|\right) + sign(y_2 - y_1) \times arctang\left(\left|\frac{y_2 - y_1}{x_2 - x_1}\right|\right) \quad \textbf{(15)}$$
$$\theta_2 = arctang\left(\left|\frac{y_3 - y_2}{x_3 - x_2}\right|\right) - sign(y_2 - y_1) \times arctang\left(\left|\frac{y_2 - y_1}{x_2 - x_1}\right|\right)$$
$$\theta_3 = 180° - (\theta_1 + \theta_2)$$

Then the score of the selected triangle, ST, is defined by Eq. (16). In Eq. (16), the errors between $\theta_{ref}(k)$ and θ_k, $_{k=1,2,3}$ are weighted by the credibilities given on the selected feature regions. Since θ_1 is based on left eye, the error between $\theta_{ref}(1)$ and θ_1 is weighted by the credibility on the feature selected as the left eye. In the same way, the error between $\theta_{ref}(2)$ and θ_2 is weighted by the credibility on the feature selected as the right eye. The credibility on an eye feature region is defined as the square of the width and height ratio. In the case of θ_3 which is based on mouth, the error between $\theta_{ref}(3)$ and θ_3 is weighted by the credibility on the feature selected as the mouth. The credibility on a mouth feature region is also defined as the square of the width and height ratio. The weighted sum of the errors of three angles is finally weighted by the ratio of the width of mouth feature region over the distance between eyes to define the score of the selected triangle. Here β is the normalization factor to make the score of non-facial region larger than 1. That is, by defining the faceness score in this way, the set of eyes and mouth features should have the score less than 1 to be selected as the final facial feature region.

$$S_T = \frac{W_3}{|x_2 - x_1| \times \beta} \times \sum_{k=1}^{3} \left(\alpha_k \times \left| \theta_k - \theta_{ref}(k) \right|^2 \right) \quad (16)$$
$$\begin{cases} \alpha_k = \left(\dfrac{H_k}{W_k}\right)^2 & for\ k = 1,2 \\ \alpha_k = \dfrac{S_3}{256} \times \left(\dfrac{H_k}{W_k}\right)^3 & for\ k = 3 \\ S_3 = \dfrac{1}{N} \times \displaystyle\sum_{l=0}^{l=N-1} (Cr_l + Cb_l) & \end{cases}$$

5 Experiments and Discussion

The proposed algorithm has been implemented using Visual C++ in the Pentium-IV PC. The performance of the algorithm is evaluated in the viewpoints of how accurately detect the face regions when the size of a face and illumination condition vary. For the experiments, more than 200 images are collected in the internet and the MIT-CMU test sets arbitrarily, so that they may have complex background, different face colors, and different face sizes.

The following Fig. 7 shows the example images and the intermediate and final results of applying the proposed algorithm.

In Fig. 7, the first column contains the original images, and the second column contains the image that is processed by the proposed method. The third column includes the candidates with the facial features, and the fourth column displays the face region (the largest rectangle) in the resulting image where the facial features are included. The proposed algorithm accurately selects the face region in the images regardless of the face size and the background brightness.

Table 1 show the comparison of face detection results between Jun Wang, et al.[10] and Proposed Algorithm from CMU database.

TABLE.1. COMPARISON OF FACE DETECTION RESULTS

	Jun Wang, et al.	**Proposed Algorithm**
Total Faces	**176**	**200**
Detected Faces	**122**	**179**
Missed Faces	**54**	**21**
Detection Rate	**69.3%**	**89.5%**

Table 1 displays that the proposed method is better performance than Wang's one with using the same CMU database.

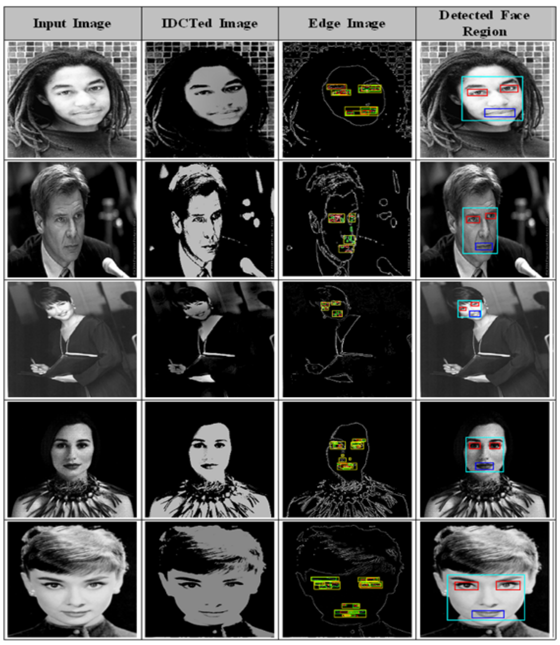


Figure 7. Example of face detection result.

6 Conclusion

This paper has proposed a new method of detecting the face regions using the characteristic of DCT that it concentrates the energy of the image on the coefficients of lower order. By removing those coefficients and taking the inverse DCT, the intensity histogram becomes to have a bimodal shape where the facial features have higher values. Therefore, the facial feature regions can be determined easily by segmenting the IDCTed image with the threshold at the center of the two modes. This effect is also found when the lighting condition is not so good that the conventional algorithms cannot find the facial features. For robust determination of DCT coefficients to be eliminated, two concepts of homomorphic filter and global contrast factor are utilized. The performance of the proposed algorithm has been tested with various face images collected arbitrarily in the networks. The experimental results have shown that the proposed algorithm accurately detects the face regions even in the images where the lighting condition and image size and orientation vary.

References:

[1] M. Yang, D. J. Kriegman, N. Ahuja : Detecting Faces in Images: *A Survey IEEE Trans. on Pattern Analysis & Machine Intelligence*, 24(1), (2002) 34-58

[2] S. Zhang and Z. Liu : A Real-Time Face Detector, 2004 IEEE International Conference on Systems, Man and Cybernetics, 3, (2004) 2197-2202

[3] G. Wei and I. K. Sethi : Omni-face detection for video/image content description, *Proceedings of the 2000 ACM workshops on Multimedia*, (2000) 185-189

[4] V. Manian and R. Vasquez : Approaches to color and texture based image classification, *SPIE JOE*, 41, (2002) 1480-1490

[5] R. Basri and D. W. Jacobs : Lambertian Reflection and Linear subspaces, *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 25(2), (2003) 218-232

[6] R. Ramamoorthi and P. Hanrahan : On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object, *Journal of the Optical Society of America*, 18(10), (2001) 2448-2459.

[7] P. Viola and M. Jones : Robust Real-Time Object Detection, *International Journal of Com-puter Vision* 57(2), 137–154, 2004.

[8] Oppenheim AV. Generalized linear filtering. In: Gold B, Rader CM, editors. *Digital Processing of Signals, New-York*, McGraw-Hill Book Company, 1969; ch. 8.

[9] K. Matkovíc, L. Neumann, A. Neumann, T. Psik, W. Purgathofer : Global Contrast Factor-a New Approach to Image Contrast, In Computational Aesthetics in Graphics, *Visualization and Imaging* 2005, pages 159-168. May 2005.

[10] J. Wang, M. S. Kankanhalli, P. Mulhem, and H. Abdulresha, “Face Detection Using DCT Coefficients in MPEG video,,”*In proceedings of International Workshop on Advanced Image Technology*, at Hualien, Taiwan, January 2002.