Real-time Live Face Detection using Face Template Matching and DCT Energy Analysis

M Hanuma Teja
BNM Institute of Technology
Bangalore, India
stellar4790@gmail.com

***Abstract*— This paper presents a method to detect the face of a live test subject in real time using a single camera as the image acquisition source. The objective is to improve the identification capability of biometric authentication systems based on face recognition by serving as a preprocessing tool to reject faces that do not belong to live humans and which can spoof the system into granting access.**

The face detection procedure involves skin area selection in the YCbCr color space and matching with a reference face template. Morphological operations such as erosion, dilation etc. have been avoided due to the uncertainty in the number of such operations that may be required under a varying real-time environment and also to reduce the computational cost of the algorithm. There are few parameters that can be considered as indicative of a live face, for example- pupil movements, nostril movements, 3D depth, and difference in number of high frequency components in live faces and photos etc. The proposed method involves the analysis of the image's DCT energy aided by detection of eye blinking and pupil movements. This yields a compact security scheme capable of resisting attacks by photographic imitation.

***Keywords*— Biometric authentication, DCT energy, Pupil movement, Face template matching, skin area threshold**

I. INTRODUCTION

The recent advances of information technology and the increasing requirement for security have resulted in a rapid development of intelligent personal identification based on biometrics. Biometrics has the capability to accurately distinguish between an authorized person and an impostor. To protect the authentication process, biometric system must be able to reject the use of a copy of a biometric instead of the live biometric. This functionality is termed "liveness detection" [1] [5].

Face recognition is one of the most active research topics due to its potential applications in access control, automated crowd surveillance, law enforcement, information safety, multimedia communication, human-machine interface, etc. Compared to other biometric authentication technologies, face recognition has an obvious advantage that it does not need much cooperation from users. Just for the broad application foreground of face recognition, how to recognize fake faces is important. A weak biometric access control can be fooled with the help of a photograph of the legitimate user. This is the problem that live face detection intends to address [5].

The possible sources of attack can be classified on the basis of what verification proof is provided to the face verification system, such as a stolen photo, stolen face photos, recorded video, 3D face models with the abilities of blinking and lip moving, amongst others[2]. To resist these attack methods, a successful live face detection system should have one or more anti-imposture abilities to expose them. Vein mapping is a very secure method of identifying a live individual, but it needs special expensive devices [3]. Other methods may require the use of multiple cameras.

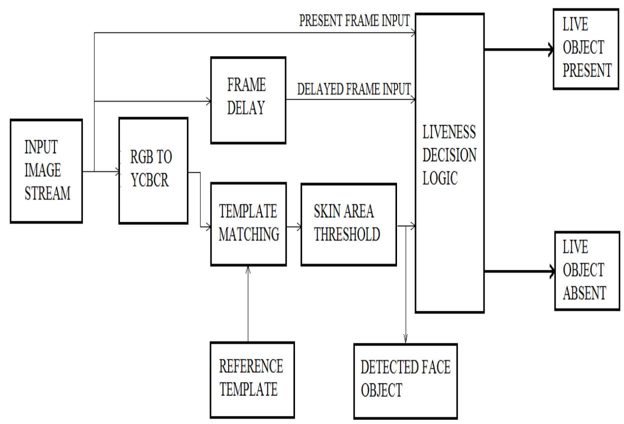

Fig.1. Block diagram of the live face detection system implemented

In the proposed method, face template matching is used for fast detection of face-like objects. The image is analysed in the YCbCr colour space and a skin area selection threshold based on the chrominance Cr of the image is then applied. In order to resist the approach of using a photo or a video recording to spoof the face recognition system, a technique based on the analysis of 2-D DCT energy is used. The algorithm is based on the fact that since the size of photo is smaller than that of live face and the photo is flat, high frequency components of photo images must be less than those of real face images. In addition, assuming that the face detected is stationary for a while, examination of residuals of subtraction between adjacent frames is used to detect pupil movements and eye blinking. This is used as a secondary parameter to establish the liveness of the face.

II. FACE DETECTION

A. Face Template Matching

This is a conceptually simple and fast method for face detection. The basic idea here is to find regions of the image which appear to be similar to the reference face template. The reference template has been generated by manually cropping and selecting the template which can best approximate a general face. Preprocessing steps that have been included are

1. Decimation of the image by a factor to reduce computational cost
2. Conversion of the input RGB images to the YCbCr color space by applying the following transformation.

$$Y' = 16 + (65.481.R' + 128.553.G' + 24.966.B')$$
$$C_B = 128 + (-37.797.R' - 74.203.G' + 112.0.B')$$
$$C_R = 128 + (112.0.R' - 93.786.G' - 18.214.B')$$

The inverse transformation is

$$R'_D = \frac{298.082.Y'}{256} + \frac{408.583.C_R}{256} - 222.921$$
$$G'_D = \frac{298.082.Y'}{256} - \frac{100.291.C_B}{256} - \frac{208.120.C_R}{256} + 135.576$$
$$B'_D = \frac{298.082.Y'}{256} + \frac{516.412.C_B}{256} - 276.836$$

This is done to bring robust performance under varying illumination conditions.

The reference face template used is shown in Fig 2.

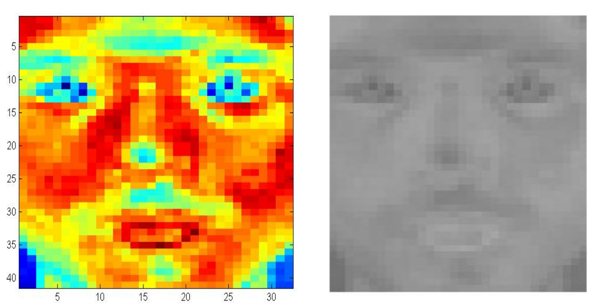

Fig.2. Chrominance component of Reference Face Template used for matching

Matching is performed by normalized 2-D cross correlation of the input image to the reference template. This is then followed by applying 2 thresholds on the value of cross correlation γ –

1. If max (γ) > 0.55 then region around maxima corresponds to a face
2. If 0.37< max (γ) <0.55 then region around maxima corresponds to a possible face object
3. If max (γ) <0.37 there is no face in the image frame

B. Skin Area Threshold

The identification of skin is performed by applying thresholds on the value of the pixel which has been converted to the YCbCr color space prior to processing. For fast classification only the chrominance Cr component is considered. A pixel is considered to be skin if its Cr component lies in the range of [132,176]. The region indicated as a possible face object by the template matching stage is now analyzed for the percentage of pixels corresponding to skin. If this value is greater than 85% then the region corresponds to a face else the object is discarded
The output after the face detection stage is shown Fig 3 & 4.

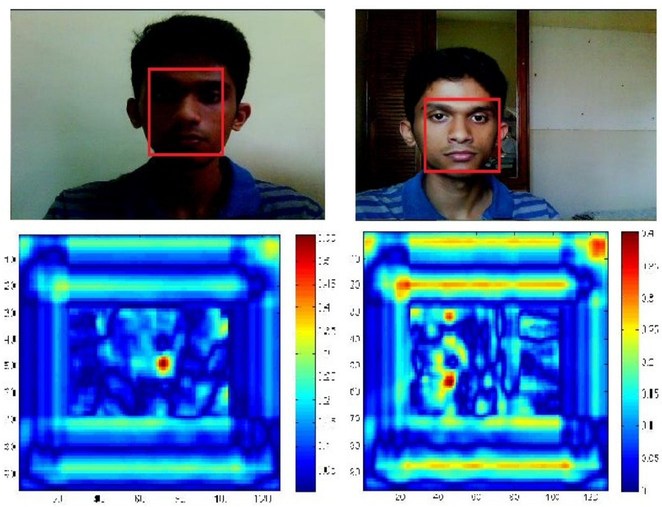

Fig.3. Face Detection output at different illumination intensities for a live face and Normalized 2-D cross correlation graphs (red areas indicate maximum correlation)

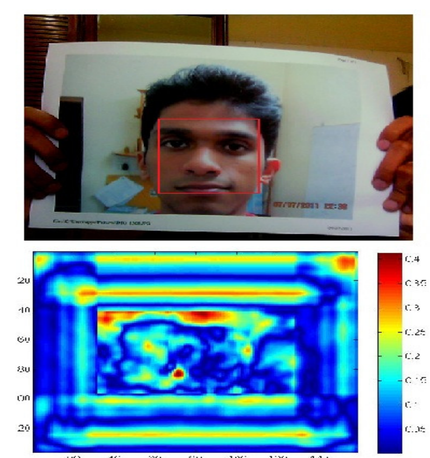

Fig.4. Face Detection output for a photographic imitation.

The correlation graphs indicate a point of maximum correlation that appears most frequently in the region of the nose. The red box around the face indicates that a face object has been detected but has yet to be confirmed as a live face.

III. LIVENESS DETECTION

The block diagram of the Liveness decision block is given in Fig 5.

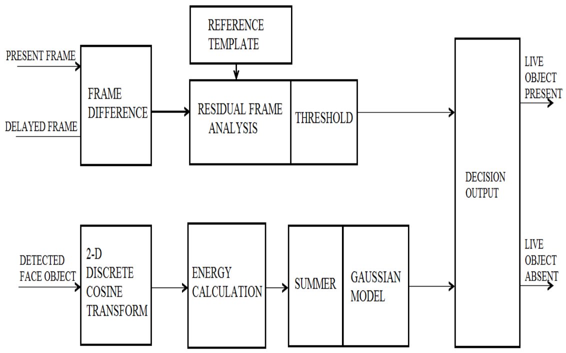

Fig.5. Block diagram of Liveness Decision Block

A. DCT Energy Analysis

This method is based on the assumption that the image of a photograph taken through a camera has a lower image quality and is therefore bound to have lesser high frequency content than the image of a live face [7]. In order to transform the image to its frequency domain representation we apply the 2-D Discrete Cosine Transform as given below-

$$F(u,v)=\left(\frac{2}{N}\right)^{1/2}\left(\frac{2}{M}\right)^{1/2}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\Lambda(i).\Lambda(j).\cos\left[\frac{\pi.u}{2.N}(2i+1)\right]\cos\left[\frac{\pi.u}{2.M}(2j+1)\right].f(i,j)$$

Where

$$\Lambda(\xi)=\begin{cases}\frac{1}{\sqrt{2}}\, for\xi=0\\ 1 otherwise\end{cases}$$

On squaring the result we get a 2-D DCT Energy Plot for each frame.

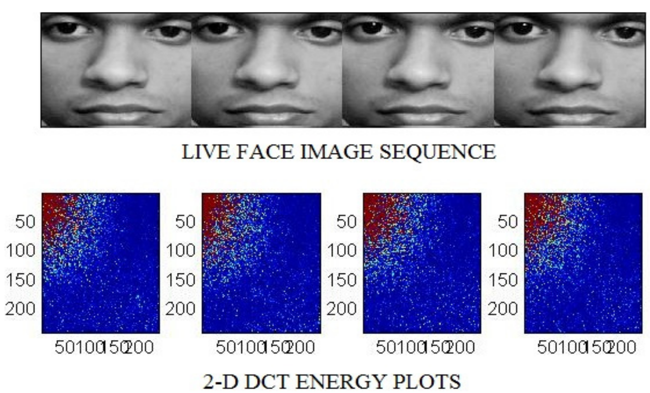

Fig.6. Image Sequence for Live Face and corresponding 2-D DCT Energy plots

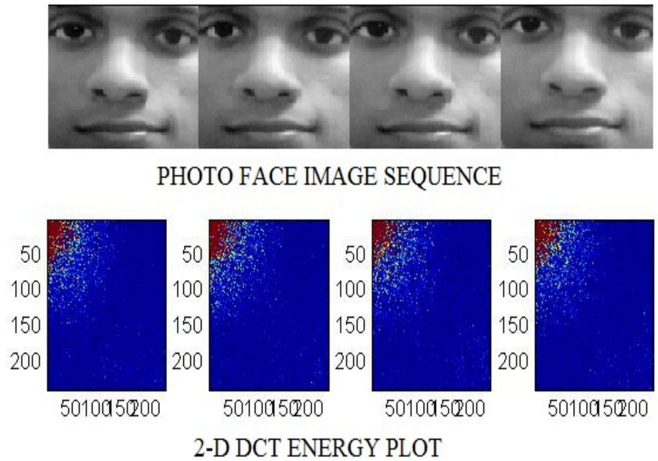

Fig.7. Image sequence for Photo Face and corresponding 2-D DCT Energy plots

The Energy plots in Fig 6 & 7 show larger high frequency content for the live face. The sum of the squared coefficients for the midscale values is used to establish a criterion for differentiating a live face and a photo. The histogram of the sum of midscale frequency coefficients calculated over 250 frames under different illumination conditions for both live and photo faces is shown in Fig 8.

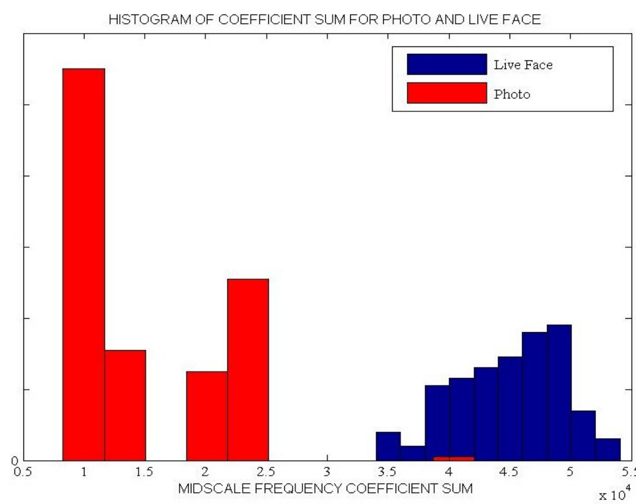

Fig.8. Histograms of sum of midscale frequency coefficients for photo and live faces

There is a very small overlap between the histograms for photos and live faces. A Gaussian distribution is defined for both sets of data. For a given frame, the coefficient sum S is calculated and the decision is made on the basis of the probability of S belonging to the live face Gaussian.

B. Detecting Pupil Movement and Blinking

The detection of pupil movements and blinking is performed using a simple technique where the adjacent frames in the detected face image stream are subtracted and the residual frame is analyzed by 2-D cross correlation with a reference eye template. The cross correlation output exhibits maxima in the region of the eyes. In order to speed up the process only the top half of the detected face is processed. The assumption made in during this process is that the face object does not exhibit sudden movement during the first few frames. A built in safeguard is added to guard against this.

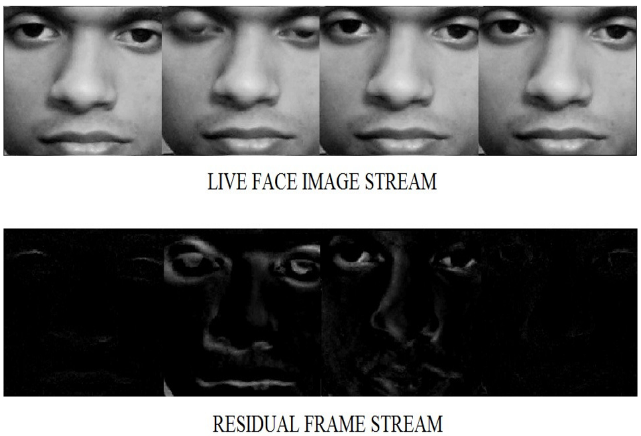

Fig.9. Example Live Face Stream and Residual Face Stream for blinking

In the face image stream shown in Fig 9, the blink occurs in frame 2. Residual frame 3 is processed to register the occurrence of the blink as in Fig 10. The threshold for normalized cross correlation in order to register a blink is set at 0.55.

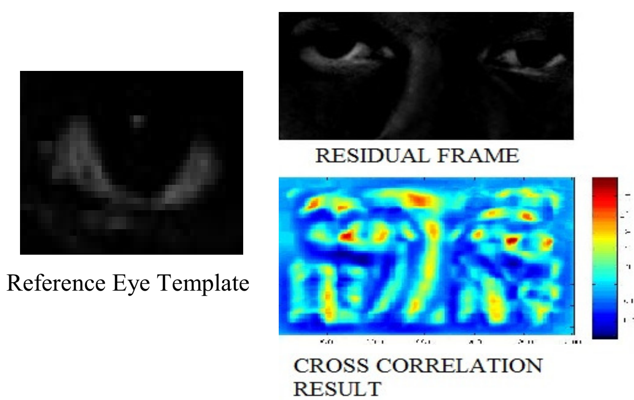

Fig.10.Reference Template and Cross Correlation output for Residual frame 3

The same technique also yields good results for detecting pupil movements as shown in Fig 11.

Robustness to sudden intensity variations or face movements can be improved by converting the residual frame for the entire input frame to a binary image and setting a threshold on the percentage area of the residual frame having luminance above a certain value. The luminance threshold for converting the residual frame to binary is taken to be 0.4 and the area must be less than 60 pixels for a valid frame. This is done only till the detected face object is determined to be live.

The residual frame stream for photo faces exhibit low correlation with the eye template when still as shown in Fig 12. Even in the case of slight motion the residual frames have maximum intensity below the valid threshold for processing.

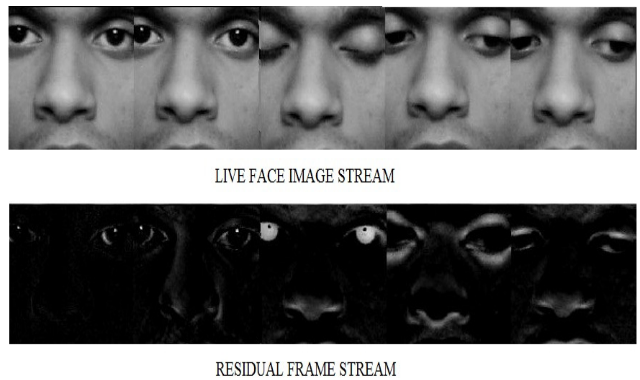

Fig.11. Example Live Face Stream for Pupil Movements

Fig.12.Example Photo Face Stream and Residual Frame Stream

The method of detecting pupil movement and blinking serves to provide a secondary parameter for selecting live faces. The disadvantage with this method is that the face of the subject must remain relatively still for at least a few frames.

IV. RESULTS

The algorithm was implemented on Matlab v7.8. 15 frames (768x1024 /frame) of the input image stream are acquired for each run. This image stream is decimated by a factor of 8 and given as input to the face detection stage. The output of 2-D cross correlation gives the approximate position of the face object. A red box indicates the detected face. The detected face image stream has a frame size of 240x240 which is given as input to the liveness decision block. The DCT Energy is found for the midscale frequencies given by the quadrant $x \in [120,240]$ and $y \in [120,240]$. The sum is fed to the Gaussian model which sets the flag for the DCT energy parameter. Next the residual frame is obtained from the original image stream and matched with the reference template. The output sets the flag for Pupil movement parameter. The sample results for both photo and live faces are shown in Fig 13 & 14.

Fig.13. Output Sequence for a Live Face. Green box indicates that the object detected is a live face

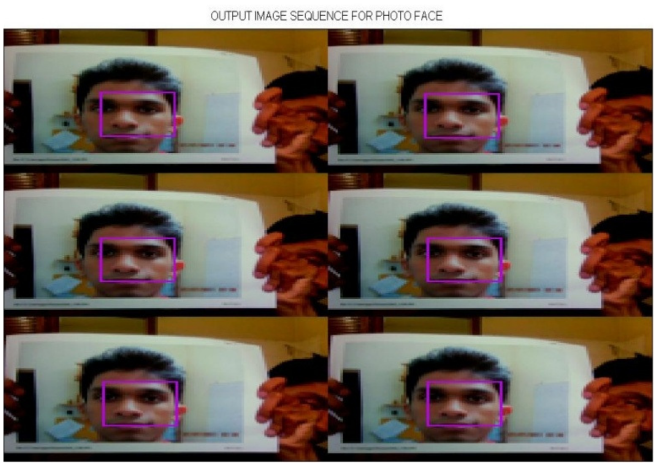

Fig.14. Output Sequence for a Photo Face. Red box persistence indicates that the object detected is not a live face.

If both flags are set the box around the detected face turns green, else it remains red. The number of frames required to determine the liveness of the face may vary from 2 to 12 frames.

V. CONCLUSION

In this paper a method was proposed to detect the face of a test subject and determine its liveness. The experimental results show that the system is able to achieve good level of discrimination between a live face and that of a photographic imitation. This was achieved with a simple algorithm without morphological operations and using only a single normal 2MP camera. Better results could be achieved with the use of multiple cameras to provide 3-D depth. The method could also be extended to live face detection for general usage in digital cameras in order to differentiate between multiple live faces and any photographs or paintings in the background.

VI. REFERENCES

[1] M K Qureshi, “Liveness Detection of Biometric Traits”, International Journal of Information Technology and Knowledge Management, January-June 2011, Volume 4, No. 1, pp. 293-295

[2], Hyung-Keun Jee, Sung-Uk Jung, and Jang-Hee Yoo, ‘‘Liveness Detection for Embedded Face Recognition System’’ International Journal of Biological and Medical Sciences 1:4 2006

[3] John R. Vacca, "Biometric technologies and verification systems", Butterworth-Heinemann, 2007

[4]Bori Toth, ‘‘Biometric Liveness Detection’’, Deloitte Information Security Bulletin, Volume 10, October 2005

[5] J. Li, Y. Wang, T. Tan, and A. K. Jain, ‘‘Live face detection based on the analysis of fourier spectra,’’ In Biometric Technology for Human Identification, SPIE vol. 5404, pp. 296-303, 2004.

[6] WY. Zhao, R. Chellappa, A. Rosenfeld, P. J. Phillips, ‘‘Face Recognition: A Literature Survey,’’ Technical Reports of Computer Vision Laboratory of University of Maryland, 2000

[7] Xiaoyang Tan, Yi Li, Jun Liu and Lin Jiang, "Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model", Computer Vision – ECCV 2010, Lecture Notes in Computer Science, 2010, Volume 6316/2010, 504-517, DOI: 10.1007/978-3-642-15567-3_37

[8] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun, Fellow, IEEE, "Real-Time Face Detection and Motion Analysis With Application in “Liveness” Assessment", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 2, NO. 3, SEPTEMBER 2007

[9] Rafael C..Gonzalez, Richard E. Woods, “Digital Image Processing”, Third Edition, Pearson Education, Inc, 2008