Chapter 6

Liveness Detection of Face Images Using 3D Fourier Transform

This chapter\(^1\) describes a frequency domain technique for live/fake face discrimination. In general, non live faces are termed as fake faces. Since the liveness of a face is mostly reflected in the high frequency of spectrum, a modified high frequency descriptor is used for proper discrimination between a fake and live face image. For generating a live video streams, the face sequences are recorded by CCD camera. For liveness discrimination between live and fake face, images are Fourier transformed in 3D and 3D frequency domain filtering is carried out in the 3D Fourier plane. The obtained preliminary results show that the task of live/fake face discrimination can be performed efficiently in such a

\(^1\)Some portions of this chapter is from the articles [122],[123],[124],[125] published and presented by the author.
6.1 Introduction

In recent years, illegally acquired images and videos containing biometric characteristics of persons are used for attacks on a system. To protect the authentication process, an automated biometric system must be able to reject the copy of a biometric instead of the live biometric. This functionality of the system is termed as *liveness detection*. For example, utilizing a fingerprint image of an administrator, an imposter may get access into the specified system using fake but authenticated fingerprint image as password. The other most common way of faking or fooling a face recognition system is to use facial photograph of a valid user to spoof the system [168]. Therefore, in practical applications, biometrics based identification systems should include the facility of correctly discriminating whether the input image is from a live subject or from a previously captured photograph or video sequences.

To resist the attacks with an intension of fooling the system with photograph, many methods are proposed in the area of face recognition [12], [13]. However, the efforts on the evolution of anti-spoofing techniques for face recognition are still very limited. The existing techniques mainly use the measurement of 3D depth and motion information, such as eye blinking, mouth movements [14]. The use of feature maps and optical flow are also extensively used [15].
6.2 Intuition of the approach

Usually, fake faces are created and exhibited in computer monitor or screen or any other form of display devices using photographs or from the pages of book / printed media. The properties of such fake faces occurring in these different media differ greatly from that of a live face as the image of a live face is found to be different from an image of a fake face. Moreover, video of a live face and video of a fake face are generally distinguishable by human eye. Human beings have the capability of distinguishing between a live face and a fake face image mainly because of 3D in nature of live face. In contrast a fake face is a 2D image.

To infer from the 3D structure of the face, an intensity model of face is required where albedo plays a vital role [98], [169], [170]. Let the image intensity be \( I(x, y) \) at a point \( P(x, y) \) on the surface of the face on the plane \( (x, y) \) and let \( A(x, y) \) be the albedo of the face at \( P(x, y) \). When light source directions are unknown a set of \( M \) images can be generated by labeling the light source directions as \( \mu = 1, \ldots, M \) and the set is given by,

\[
I(x, y) = A(x, y)N(x)S(\mu) \tag{6.1}
\]

where \( N(x) \) is its surface normal (3D shape) to the face at \( x \), and \( S(\mu) \) is an arbitrary light source direction.

Equation 6.1 can be computed for finding albedo, shape and light source directions [171]. Some general observations on the solution can be mentioned as:

1. Due to 2D planar structure of photograph, \( N(x) \) is a constant and therefore,
under the same illumination, images from a live face are determined by the albedo and the surface normal, whereas those from a fake recoded from photograph are determined by the albedo only.

2. The intensity contrast of live face image is more obvious than that of fake image. Such differences led to their greatly different reflectivity of light.

3. In most cases, the size of the fake image is small, therefore it does not contain small details of live face.

4. The frequency spectrum of the fake face (face photo) and live face are different. A live face has more high frequency content, simply because the live face can never be absolutely still in the true sense of the term. The differences particularly at higher frequencies can produce a discriminating criterion between live and fake faces.

6.3 Fourier domain approaches

To resist the approach of using a photo to spoof the face recognition system, a technique based on the analysis of 2D Fourier spectra is proposed [169], as the frequency distributions of live and fake faces are differently reflected. 2D Fourier transform \( F(u, v) \) of an image \( f(x, y) \) is given by,

\[
F(u, v) = \sum_{y=0}^{N_y-1} \sum_{x=0}^{N_x-1} f(x, y) \times \exp[-2j\pi \left( \frac{ux}{N_x} + \frac{vy}{N_y} \right)] \tag{6.2}
\]

Extracted magnitude of Fourier spectra of the live and fake face are used to
calculate the frequency content of certain area of the total spectra and a high frequency descriptor of the region of interest is defined. The high frequency descriptor (HFD) is calculated as the ratio of the energy of high frequency components to that of all frequency components at the Fourier plane and is given by [169],

\[
HFD = \frac{\int \int_{f > 2/3 f_{\text{max}}} |F(u, v)| du dv}{\int \int |F(u, v)| du dv - F(0, 0)}
\]  

(6.3)

where \(F(u, v)\) is the Fourier transform of the face image, \(f_{\text{max}}\) is the highest radius encompassing the spatial frequencies of \(F(u, v)\).

The amplitude of the spectrum at \((0, 0)\) given by \(F(0, 0)\), carries no information and hence can be filtered out when HFD is calculated. It is also heuristic to take the upper frequency limit to 2/3 of maximum frequency that may be present in the frequency spectrum. Since the distribution of frequencies near \((0,0)\) is almost a Gaussian, it is not justifiable to deduct only \(F(0, 0)\) in the denominator of equation 6.3, as proposed in [169]. Hence, the question of selecting the frequency band is also open. Therefore a modified high frequency descriptor (MHFD) is proposed by the author as [123],

\[
MHFD = \frac{\int \int_{f_{\text{min}} < f < f_{\text{max}}} |F(u, v)| du dv}{\int \int |F(u, v)| du dv - \int \int_{f_{\text{min}}} |F(u, v)| du dv}
\]  

(6.4)

As evident, the MHDF does not require any amplitude thresholding. Only the selection of frequency band between maximum and minimum frequency is required. This band of frequency needs to be calculated to discriminate between
live face and a fake face image. However, it is difficult to set minimum frequency limit, even in the case of modified high frequency descriptor. Moreover, the method proposed has little use in case of video strips of fake and live images. Therefore, an alternative approach of using 3D Fourier transformed information is proposed for liveness detection of video face images. The filtering of the frequency region is done by a upper and lower limits over several images of the time frames of the video.

6.4 Brief description of 3D fast Fourier transform (3D-FFT) technique

The usual method of 2D Fourier Transform of an image is done on the pixels in 2D plane. 3D FFT [172], can be briefly stated as obtaining FFT of the voxels in three dimensional plane. The voxels can be generated when the frames of a video sequence are used.

Let \( f_t(x,y) \) be a 2D gray image of size \( N_x \times N_y \) at a time frame \( t \). A sequence of \( N_t \) images \( f_t(x,y) \) can be represented as a 3D image \( f(x,y,t) \), where \( x = 0,1,2,...,(N_x-1) ; y = 0,1,2,...,(N_y-1) \), and \( t = 0,1,2,...,(N_t-1) \). The discrete Fourier transform of a 2D spatio-temporal images \( f(x,y,t) \) of size \( N_x \times N_y \times N_t \) in time frames is given by,

\[
F(u,v,w) = \sum_{t=0}^{N_t-1} \sum_{y=0}^{N_y-1} \sum_{x=0}^{N_x-1} f(x,y,t) \times \exp[-2j\pi(\frac{ux}{N_x} + \frac{vy}{N_y} + \frac{wt}{N_t})] \tag{6.5}
\]
where, \( u = 0, 1, 2, ..., (N_x - 1); v = 0, 1, 2, ..., (N_y - 1) \) and \( w = 0, 1, 2, ..., (N_t - 1) \).

3D DFT computation can be efficiently performed by the row-column-frame decomposition algorithm [173]. It is based on a decomposition of the 3D DFTs in \( N_y \times N_z \) 1D DFTs along the row, \( N_x \times N_t \) 1D DFTs along the column and \( N_x \times N_y \) 1D DFTs along the frame of a video sequence as given by,

\[
F'(u, v, w) = \sum_{x=0}^{N_x-1} f(x, y, t) \times \exp[-2j\pi\left(\frac{ux}{N_x}\right)] \tag{6.8}
\]

\[
F''(u, v, t) = \sum_{y=0}^{N_y-1} F'(u, y, t) \times \exp[-2j\pi\left(\frac{vy}{N_y}\right)] \tag{6.7}
\]

\[
F(u, v, w) = \sum_{t=0}^{N_t-1} F''(u, v, t) \times \exp[-2j\pi\left(\frac{wt}{N_t}\right)] \tag{6.8}
\]

Computations of 3D FFT can now be performed with the help of these three equations shown above [174].

### 6.5 Proposed process sequence of the live/fake discrimination

The technique proposed for live / fake discrimination using 3D FFT is given by the following steps:

**Step 1:**
Take the input live face image sequence recorded by a CCD camera. Select the input frames with a gap of 10 frames and obtain the 3D Fourier transform of the selected input frame sequence of the live images.

**Step 2:**

Generate a fake video sequence from a photograph of the same person and obtain 3D FFT of the sequence.

**Step 3:**

Select a hollow cylinder with inner and outer radius $r_{\text{min}}$ to $r_{\text{max}}$ and calculate the average frequency content of that region for both live and videos.

**Step 4:**

A decision threshold is set where the energy of the frequency difference is high to arrive at a decision whether the recorded sequences is a live face or generated from a fake video sequence of a photographs.

### 6.5.1 Calculation of frequency content of the hollow cylindrical region on 3D Fourier domain

Frequency content of fake and live images is obtained from the 3D Fourier transform of the video sequence. However, it is necessary to quantify the frequency content in terms of energy for distinguishing live and fake images. To calculate the energy of frequency components within the radii $r_{\text{min}}$ and $r_{\text{max}}$, a band reject filter is constructed by eliminating the frequencies outside the $r_{\text{min}}$ and $r_{\text{max}}$ as shown in Figure 6.1. The average energy within the hollow cylinder
is calculated as,

\[ E(r) = \frac{1}{n} \sum_r P(u, v, w) \]  

(6.9)

where, \( n \) is the number of frames taken.

It still remains heuristic to find the lower and upper frequency limits that may be present in the face sequences. To make the system more robust, the average energy of the frequency content of the hollow region versus annular radius is plotted as shown in Figure 6.1. Then a suitable threshold for discrimination is selected where the discrimination is maximum.

Figure 6.1: Selection of the hollow ring
6.5.2 Experiments

To establish the validity of the proposed 3D Fourier transform technique for liveness detection, the live videos of 40 different persons (30 males and 10 females) are recorded with a CCD camera. These live videos naturally contain eye movements, small changes in muscle movements and other changes in face sequences. For example, when a person is in joyous mood, the movement of eye brows, locations of iris, twitching of nose, muscle movements in cheeks and mouth portions may likely be present in the live videos. In contrast, the expression and pose of the fake faces are invariant to such small changes. A subset of 20 image-frames are constructed from live videos by extracting a frame for every 10 continuous frames.

The image sequences are shown in Figure 6.2. For every person, a fake image sequence of the face is generated from a particular photograph of that person captured by the same video camera as shown in Figure 6.3 named as fake video 2. Another set of images as shown in Figure 6.4 (named as fake video 1) of the same photograph is captured with a little shift in left and right.

Figure 6.5 shows a sequence of 3D FFT of the live face images of Figure 6.2. Figure 6.6 6.7 show a sequence of 3D FFT of the fake face photograph sequences and shifted video sequences of Figure 6.3 and Figure 6.4.

Figure 6.2: A live video image sequence used for experimentation
Chapter 6: Liveness Detection of Face Images Using 3D Fourier Transform

Figure 6.3: Fake image (photograph) sequence of the same person

Figure 6.4: Fake image sequence (with left and right shifted) of the same person

Figure 6.5: 3D FFT of live image sequence of the same person

Figure 6.6: 3D FFT of Fake image (photograph) sequence of the same person

Figure 6.7: 3D FFT of Fake image (photograph) sequence (with left and right shifted) of the same person
6.5.2.1 Adjustment of the highest and lowest frequency

The difference between the live face and fake face in frequency domain is evident from the Figures. Sharp zero order in the fake face spectrum gives way to the Gaussian distribution. Because of the motions and changes in posture in live face more higher order frequencies are seen in the spectra of live face sequences.

A plot of the energy of the frequency contents is obtained by varying $r$ from center to the highest frequency as shown in Figure 6.8. For each case, the $E(r) - r$ curve for the live faces contains more energy value (because of more high frequency contents) than the fake faces as shown in blue line with star. It can be observed from the plot that fake video 2 has no change in frequency content as shown in red line, whereas for fake video 1, variation in frequency content (shown in blue line) is observed. The change in high frequency content in video set 1 is due to
the positional change in the photograph. However, the change of energy content of live face is always high as shown in the plot. The energy difference is highest in the region where \( r \) lies between \( r_{\text{max}} = 35 \) to \( r_{\text{min}} = 32 \). This sets the range of radius values which can be considered for discrimination of live and face videos. The experiment is carried out on 40 persons and the results using the proposed method gave 100 percent accuracy in each case.

### 6.6 Discussions and conclusions

This chapter introduces exploratory results for live face discrimination technique in frequency domain, where 2D and 3D FFT are computed over a set of face frames recorded by a CCD camera. High frequency discriminator is used to exhibit the biometric characteristics of live face. In may be concluded that 3D FFT is suitable for processing the video frames. The measured energy content due frequencies in the hollow cylinder in the frequency domain defines effectively the decision threshold in preventing spoofing the system with photograph of the subject.