Identification of 2-D facial Spoofing attacks using Texture and Gaussian Mixture Model Based Fourier Analysis

Ayush Rai, Tarun Krishna, Shubham Gupta, Shubham Bansal, Shubham Khandelwal, Sonam Nahar
The LNM Institute of Information Technology, Jaipur, Rajasthan, India
{ayushraiformula1, krishnatarun7, shubhamgupta5893, shubbansal27, skhlnmiit, sonamnahar}@gmail.com

Dushyant Goyal
Stony Brook University, New York, USA
goyal1dushyant@gmail.com

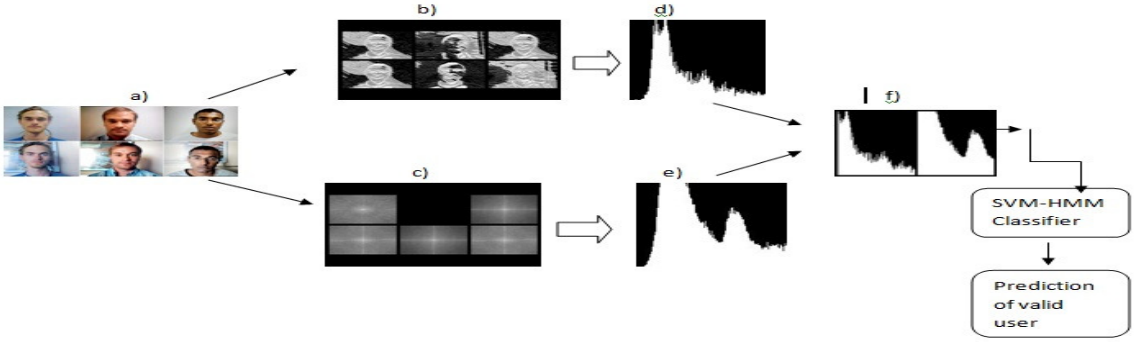

Figure 1. a)Input Image having both types of persons i.e. Real and Fake b) Output of LBP c) Output of GMM_Fourier d) Histogram of LBP e) Histogram of GMM_Fourier f) Concatenation of average histogram, Further we use SVM-HMM hybrid model for the classification

ABSTRACT

Continuous development in the field of Biometrics has definitely contributed much to enhance the efficiency of the 2-D Face Authentication Systems. But, despite such progress, such Authentication systems are still susceptible to pseudo attacks, where users without actual access tend to authenticate themselves as valid users. In our work, we present a solution to video based spoofing attacks to face biometric systems. The unique thing about the suggested approach is that it not only works on printed and mobile attacks but also on high definition video attacks. The methodology adopted by us deals with two major ideas to distinguish between real and fake person. Firstly, we exploit the texture feature of the attack in spatial domain through Local Binary Pattern (LBP). The Second idea involves taking into account the noise generated by the recaptured video through Background Modeling by Gaussian Mixture Model (GMM) followed by the Discrete Fourier Transform of the video. The results obtained by this approach on REPLAY-ATTACK database have produced excellent results, which show the potency of the adopted strategy.

***Index Terms*—** Local Binary Pattern (LBP), Gaussian Mixture Model (GMM), Discrete Fourier Transform (DFT), Video Based Spoofing.

1. INTRODUCTION

The problem of the spoofing attacks has become an active research area during the last few years especially when we need the high security. These attacks are generally performed by the invalid user who tries to gain an access to the security system by presenting the photo or the video of the valid person in front of the camera i.e. by presenting a copy of a valid user through some electronic media. The images of a valid user can easily be searched on the internet. To solve this kind of problem, many of the state-of-the-art algorithms have been evaluated on databases NUAA [1], Print-Attack and CASIA-FASD [2], REPLAY-ATTACK [3].Motion-based algorithm exploits correlations between the client head movements and the scene context in the case of spoofing attacks. To capture the temporal information contained in the Fourier spectrum which is called visual rhythm is proposed in [4]. The textural features are used from Gray-Level Co-occurrence Matrix (GLCM) in [5]. Use of Gabor wavelet features for texture analysis and the feature vector [6] is made by using mean and standard deviation of the magnitude of the transform coefficients. Gaussian mixture model (GMM) for background modeling [7] is also used. In this, each pixel is classified based on whether the Gaussian distribution, which represents it most effectively, is considered a part of the background model. Video and static analysis is done in [8] in order to get the information about motion, texture and liveness. The Analysis of the quality degradation of the attack samples by capturing textural information [9] is done using multi-scale local binary patterns (LBP). Gray-scale and LBP feature is computed [10] using rotation invariant uniform local binary patterns. It takes a standard video sequence as input, and applies spatial decomposition by 2 main approaches: Laplacian and Gaussian pyramids. Extraction of face background consistency feature is done in [11] which is also known as complementary non-rigid motion analysis approach. Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) operator [12] is used for combining both space and time information into a single operator. REPLAY-ATTACK database [3], a novel spoofing attack database, has motivated the 2nd Competition on Counter Measures to 2D Face Spoofing Attacks.Our algorithm comprises of the algorithms like LBP features similar to [9] and non-rigid motion analysis approach, for which the face background consistency feature [11] and Gaussian Mixture Model (GMM) [9] are also used. Our work was done on the dataset provided for the 2nd competition on counter measures to 2D facial spoofing attacks ICB 2013 [3].

The remaining paper is organized as follows. Sections 2 describes about the database which we used in our work. Section 3 proposes the LBP-GMM-Fourier based approach for the spoofing detection. In Section 4, feature extraction and classification algorithms used are described. Section 5 includes Experimental Analysis while Section 6 comprises of future approach for spoofing detection using kinect sensor and Section 7 contains conclusion.

2. REPLAY ATTACK DATABASE

The Replay attack Database consists of video clips of photo and video attack attempts of 50 clients, under different lighting conditions. Real client accesses as well as data collected for the attacks are taken under two different lighting conditions: (1) Controlled (the background is homogeneous); and (2) adverse (the background is non-homogeneous). With regards to the support the attack media, an attack as described in one of the following scenarios: (1) print (produced on a Triumph-Adler DCC 2520 color laser printer occupying the whole available printing surface on A4 paper); (2) mobile (videos are taken using an iPhone 3GS screen with resolution 480x320 pixels); (3) highdef (videos are taken using an iPad with a screen resolution of 1024x768 pixels). The attacks are divided into two different types of attack modes: (1) fixed-support attacks (It is composed of videos generated using a stand to hold the client biometry). (2) hand-based attacks (the attacker holds the device used for the attack with their own hands). This database has several advantages compare to other face spoofing databases such as NUAA, CASIA-FASD. The videos in the REPLAY-ATTACK database consist of 3 subsets i.e. for training, development and testing. Training set to be used for training the anti-spoof classifier, Development set to find the threshold estimation, Test set which is generally used to report the results and error figures. An important thing to notice is that every set among Train, Development and Test is comprised of real access and three types of attacks namely Printed attacks, Phone attacks and Tablet attacks.

3. PROPOSED ALGORITHM

In this work, we present a different approach to tackle the spoofing problems in the world today. Local Binary Pattern (LBP) utilizes the poor texture quality of the printed photographs. Therefore, it is one of the superior techniques for Anti-spoofing 2-D Facial Recognition System. But now, we not only encounter photographic printed attacks in real world. The suggested strategy is meant to detect High-Definition video attacks. The spatial domain of the image does not provide us the enough information about the image. Hence, we take into account the Fourier domain of the video. We introduce the new feature, which we call as GMM_Fourier. GMM_Fourier is basically a combination of Background Modeling through Gaussian Mixture Model followed by Discrete Fourier Transform (DFT) of the samples.

3.1. Overview of Local Binary Pattern

LBP is the type of technique which generally performs the texture analysis of the images. It can be defined as a gray scale invariant texture measure [10]. This is based on comparing the value of the central pixel with the value of its neighboring pixels. Detailed description is shown in the figure 2. We applied this operator not only on the face but, on the whole image.

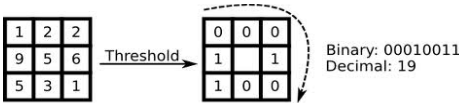

Figure 2. The basic LBP operator

The LBP Operator can be denoted as the Expression $LBP_{X,Y}$, where X represents the total number of sampling points in the neighborhood which we are considering for application of LBP Operator and Y denotes the radius of LBP Operator. The LBP Code for a pixel (I_c ,J_c) is calculated through the expression:

$$LBP_{x,y} = \sum_{x=0}^{X-1} Z\left(G_p - G_c\right)\left(2^x\right) \quad (1)$$

where, G_p denotes the grayscale value of the centre pixel and G_c denotes the grayscale value of the sample neighbor pixel. Z is a function which is taken into account for the purpose of thresholding with following description.

$$Z(n) = \begin{cases} 1; & if\ n \geq 0 \\ 0; & otherwise \end{cases} \qquad (2)$$

The output of this technique, clearly differentiate between the real and fake faces as seen in figure 3. In order to extract features from the images, we make use of the histogram.

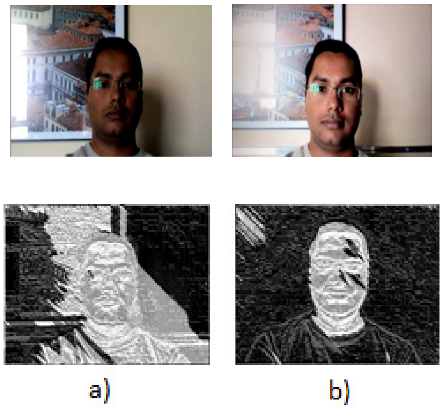

Figure 3. a) Real Person with its LBP Output b) Fake Person with its LBP Ouput.

3.2. Spoofing Detection based on GMM with Fourier

To solve the problem, we combined two existing methods. The first being GMM which is one of the methods of background modeling which will be useful in describing the motion in scene. Further, we try to differentiate on the basis of noise that is dealing the frames in frequency domain. A GMM can be defined as a parametric probability density function as a weighted sum of Gaussian Component densities. Let the conditional density for a pixel C belonging to a multi colored object A be a mixture with M component densities.

$$p(C|A) = \sum_{i=1}^{M} p(C|i)P(i) \qquad (3)$$

Where a $P(i)$ is called a mixing parameter, which corresponds to the prior probability that pixel A was generated by component i and also $\sum_{i=1}^{M} P(i) = 1$. Each Mixture component is a Gaussian with mean μ and covariance matrix $\sum$ i.e. in the case of 2D Color Space.

$$p(C|i) = \frac{1}{2\pi|\Sigma_i|} \exp^{-1/2(C-\mu i)^\wedge T \Sigma_j^{-1}(C-\mu i)} \qquad (4)$$

We use GMM to describe the scenic motions of faces. By this, we can differentiate between foreground and background regions. After applying GMM on each frame, we now try to look into frequency spectrum of the obtained frames.

We can detect the noise which is contained in the video by applying a 2D discrete Fourier transform on each frame of the noisy video.

For a square size Image i.e. N×N DFT is given by:

$$F(x,y) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(n,m) e^{-i2\pi\left(\frac{xn}{N}+\frac{ym}{N}\right)} \qquad (5)$$

But, we do not have our frames of square size N×M, so in this case, this turn up into:

$$F(x,y) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f(n,m) e^{-i2\pi\left(\frac{xn}{N}+\frac{ym}{N}\right)} \qquad (6)$$

The figure 4 shown below shows such differences between the fake and real user.

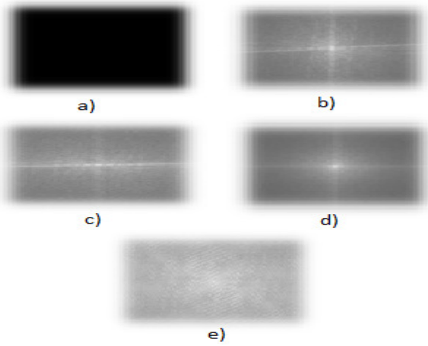

Figure 4. a) Fixed photo b) Fixed video c) Hand Video d) Hand photo e) Real Person.

$$|F(x,y)| = \sqrt{I(x,y)^2 + R(x,y)^2}$$
$$H(x,y) = \log(1 + |F(x,y)|) \qquad (7)$$

Figure 4.a – 4.d, show the output of fake persons and 4.e shows the output of the real person. Observing each frame, one can easily differentiate between the real face and fake face.

4. FEATURE EXTRACTION AND CLASSIFICATION

Let us consider a video sample V, which consists of N number of frames. We apply LBP on every frame of the training sample and hence we compute an Average histogram of all images of the video. We use 256 numbers of bins here. Therefore, our feature vector *F(a)* contains 256 features. Further, we also compute the average histogram of all the frames obtained after the application of the GMM_Fourier Technique. Again, we use 256 numbers of bins here. So our feature vector *F(b)* consists of another 256 features. Now, we merge both the features by concatenation of their histograms. As a result, our final feature vector consists of *F(a)+F(b)=512* features. In this paper, we use the hybrid SVM-HMM method proposed in [13] a publicly available toolkit which takes advantage of inherent properties of both SVM and HMM, to train the SVM-HMM Classifier by making use of these features. In the Similar Manner, we extract same features from the development/test sample and use SVM_HMM Classifier to perform the prediction.

5. EXPERIMENTAL ANALYSIS

For training and classification, we have used Svm_Hmm hybrid Model. Since the temporal dynamics of facial action can be described very effectively by HMM's. The classification of various facial features in every frame is generally achieved using Gaussian Mixture Models as the emission probabilities. The Gaussian Mixtures are trained by likelihood maximization, which assumes correctness of the models and thus suffers from poor discrimination of real and fake person while SVM's on the other hand discriminate extremely well.

Table 1: Results obtained on Development Set and Test Set

Method	Type of Attacks	Devel. Set Accuracy %	Test Set Accuracy %
LBP (256 features) [3]	Tablet (High Def.)	96.52	96.94
	Phone	97.8	98.28
	Print	99.2	99.34
GMM-Fourier (256 features) [11],[14]			
	Tablet (High Def.)	99.44	98.5
	Phone	97.26	96.42
	Print	99.00	97.8
LBP + GMM-Fourier (512 features)			
	Tablet (High Def.)	100	100
	Phone	100	100
	Print	100	100

Therefore with help of their mixtures we can have an hybrid model i.e. Svm_Hmm to discriminate real and fake person from each other. The use of such a hybrid model has helped to increase our accuracy of classification**.** We used SVM_HMM Classifier to train our model and then classify the videos in the test/development set into real or fake category. The comparison of accuracy achieved by Svm_Hmm using different features i.e. LBP, GMM-Fourier and their fusion on Development and Test Set has been depicted in Table 1. The overall accuracy obtained on Development Set by using the combination LBP and GMM_Fourier was 100%. Similarly on Test Set we obtained the accuracy 100%.

6. FUTURE WORK

By using Kinect Sensor Cameras, we can easily distinguish between a real person and his imposter attack based on their depth information. Through Kinect Cameras depth estimation of the user is done. By observation of the depth of foreground and background in the frames, spoofing detection is done. In case of real access the foreground and background would have different depths while in case of an attack the depths of foreground and background remains the same. This difference can be easily observed in the figure 5. We are not proposing this approach now rather we only want to mention that such methods would be vital in future.

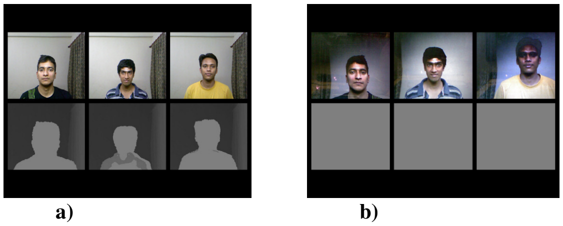

Figure 5.a) Real Subjects with their depth images b)Attacks with their depth images.

7. CONCLUSION

Video-Based Spoofing attacks are more challenging as in case of video, methods like blink detection and motion detection often fails.. In this paper, we present a method for detection of video based spoofing attacks particularly for High Definition video by targeting the texture of the various attacks through Local Binary Patterns (LBP) and the noise generated by recaptured video through Background Modeling followed by DFT of the video. Our technique is more reliable as it works on high definition videos also.

8. REFERENCES

[1] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In ECCV (6), pages 504–517, 2010.

[2] Z. Zhiwei, Y. Junjie, L. Sifei, L. Zhen, Y. Dong, and S. Z. Li. A face anti spoofing database with diverse attacks. In Proceedings of the 5th IAPR International Conference on Biometrics (ICB'12), New Delhi, India, 2012.

[3] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing.

In Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the, pages 1–7, 2012.

[4] S. Guimaraes, M. Couprie, N. Leite, and A. Araujo. A Method for Cut Detection Based on Visual Rhythm. In SIBGRAPI, pages 297– 304, 2001.

[5] R. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. IEEE TSMC, SMC-3(6):610 –621, nov. 1973.

[6] B. Manjunath andW. Ma. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8):837–842, 1996.

[7] Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999.

[8] R. Tronci et al. Fusion of multiple clues for photo-attack detection in face recognition systems. In IJCB, pages 1–6, 2011.

[9] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In International Joint Conference on Biometrics, pages 1–7, 2011.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971 –987, July 2002.

[11] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Li. Face Livenes detection by exploring multiple scenic clues, In 12th International Conference on Control, Automation, Robotics and Vision, 2012.

[12] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. LBP-TOP based countermeasure against face spoofing attacks. In International Workshop on Computer Vision With Local Binary Pattern Variants - ACCV, page 12, 2012.

[13] Y. Altun, I. Tsochantaridis, and T. Hofmann. Hidden markov support vector machines. In International Conference on Machine Learning, 2003.

[14] Allan da ,Silva Pinto, Helio Pedrini, William Robson Schwartz, Anderson Rocha. Video-Based Face Spoofing Detection through Visual Rhythm Analysis.