Face Spoof Detection Using VGG-Face Architecture

Balamurali K 1, Chandru S 2, Muhammed Sohail Razvi 3 and V. Sathiesh Kumar 4

1 2 3 4 Department of Electronics Engineering, MIT Campus, Anna University, Tamilnadu, India.

Email: k.balamurali303@gmail.com

Abstract. Face recognition systems have been obtaining substantial importance in modern world. Security systems are major application of face recognition system. However, the potential of the face recognition system to withstand the attack of an unauthorized person is an important concern. Face recognition systems are vulnerable to photographs and video spoof attacks. In these scenarios, anti-spoofing systems comes in handy to evade these attacks. Robust solutions are required for face recognition system to be immune against spoofing attacks. In this paper, the detected face is denoised and then converted to YCbCr and CIELUV colour model and then passed through VGG-Face architecture for extraction of face embeddings of each colour space. Then the extracted face embeddings are concatenated and then passed through SVC (Support Vector Classifier) which then classifies real and spoof faces. The proposed method has obtained a test accuracy of 99.6% with specificity of 99.5% for spoof detection.

1. Introduction

Biometric systems such as face recognition, finger-print identification is extensively used for personal identification. It is more secured than any traditional methods like passcode-entry, ID card, or keys. Face recognition system is also more convenient than the traditional methods. However, Face recognition is often prone to presentation attacks. Presentation attack includes print and video/replay attacks. In print attack, the attacker utilizes the photo of a valid user presented in a digital device or printed in a paper. In video/replay attack, the attacker uses natural human movements of a valid user recorded in a video. Many different types of hardware and software methods have been developed to detect spoof faces.

The software-based methods analyse the liveness properties such as textures, structure information and liveness sign and image quality. All these methods are very sensitive to environment noises such as low light conditions. It takes high computational time to obtain the result. Texture based methods utilizes the reflection of light from the surface of the target. The human skin will reflect the light differently than a plain paper or a screen. Detection of spoof attack is based on the difference between real and spoof face’s visual and tactile quality. Tactile texture represents the roughness or smoothness of the surface while visual texture is the illusion of quality of the surface. The algorithms used for the texture-based methods are local binary pattern (LBP), Fourier analysis, Colour Texture analysis. This method is more subjective to the environmental noises. Structure based methods captures the differences in the structural properties of 2-D plain surface and 3-D surfaces. The light falls on the 3-D surfaces diffuses more slowly than the 2-D surface. The delay in the diffusion is due to nonuniform light spread on the 3-D surface. It requires more time for detection. The liveness
detection method utilizes blinking of eyes and movement of mouth for spoof detection. It also requires
the cooperation of the users. This type of spoof detection fails when it comes to video/replay attacks.
The image quality analysis method detects the quality of the image used for detection. Higher
resolution images can pass these kinds of analysis methods. Hardware based methods uses additional
hardware such as infra-red cameras, multiple 2-D cameras provide high accuracy results but it costs
high. In this paper, the proposed system performs spoof detection computation by combining the
features extracted from CIELUV and YCbCr colour space converted detected and denoised face.

2. Literature Review
Yousef et al., have utilized patch-based CNN for extraction of local features and depth-based CNN for
depth map generation which then used to identify the real and spoof face images. Patch-based CNN
has been used for increasing the training data and to retain the native resolution of original image.
Classification is done using the features extracted from the depth map using depth-based CNN.
Finally, fusion analysis has been done for spoof detection achieving an EER of 0.35±0.19 and HTER
of 0.21±0.21 on MSU-USSA dataset [1]. Lei et al., have come up with a method involving DPCNN to
eXtract the part features of convolutional layers. The dimensions of output of third convolutional layer
of fine-tuned VGG-Face have been reduced using PCA (Principal Component Analysis) which then
advanced through SVC to classify the real and fake images achieving an EER of 5.0 and 4.5 in CASIA
database [2].

Zinelabidine et al., achieved an EER of 6.2% on CASIA database and 0.4% of EER and 2.9% of
HTER on replay attack database by concatenating the histograms of each colour channel in YCbCr
converted image [3]. Shatish et al., developed an approach of converting the image to L*a*b* colour
space which then passed through HOG (Histogram of Oriented Gradients) for face detection. The
detected face is then advanced through VGG7 CNN architecture for extraction of features and then
classified using fully connected layer for spoof detection. They were able to classify 1949 images
correctly out of 2135 with an accuracy of 91.29% [4].

Jukka et al., has proposed an approach which utilizes multi-scale LBP for texture analysis of
facial images and encodes the micro-texture into an enhanced histogram. Then a SVC (Support Vector
Classifier) is utilized to determine spoof faces. The proposed method has achieved an accuracy of
98%, FAR of 0.6% and FRR of 4.4% [5]. Kilioglu et al., has put forward a method for liveness
detection built on the pupil movement achieving a success ratio of 89.7% and 94.8% on database that
excluded persons with glasses [6].

Kant et al., proposed an approach utilizing both camera and thermal sensor. For detection both
the camera and thermal sensor captures the users and then for any frame it compared with thermal
image from the thermal sensor which discriminates the face skin from the 2-D surface. They have
achieved an accuracy of 98% using thermal face recognition [7]. Jukka et al., has proposed a method
involving the passage of input image through face detector and upper body detector. Upon finding the
upper body, then the image is passed through spoofing medium detector for further classification of
real and spoof faces. Combining both the CASIA and NUAA dataset they have achieved an EER of
6.8% [8].

3. Dataset
The dataset utilized for training and testing of face spoof detection consists of real and spoof images
from NUAA photography imposter dataset [8] and custom created dataset. The number of real and
spoof images with respect to NUAA photography imposter dataset and custom created dataset are
shown in Table 1. The samples of respective datasets are shown in Figure 1&2.

| CLASS       | NUAA DATASET | CUSTOM DATSET |
|-------------|--------------|---------------|
| REAL        | 3781         | 919           |
| SPOOF       | 2398         | 2371          |
4. Methodology
In this section, the working of facial anti-spoofing system is discussed here. Facial anti-spoofing is the task of averting presentation attack for false facial authentication. The proposed system in this paper is illustrated in Figure 3.

In this paper, face detection has been performed using MTCNN (Multi-Task Cascaded Convolutional Neural Network) [10]. MTCNN comprises of three networks corresponding to each step respectively. Initially, an input image is passed through P-net for prediction of possible face positions and their bounding boxes. The respective output consists of large number of false positives. Hence the output is passed through R-net for regression of bounding boxes which eliminate false positives.
positive and improves accuracy. Further the output of R-net is passed through O-net for refinement of bounding boxes. The detected face is then denoised using Non-Local means denoising algorithm which replaces the target pixel based on mean value of all pixels in an image and similarity of all pixels with target pixel. Then the denoised face is then converted into YCbCr and CIELUV colour space. In YCbCr, Y represents the luma component of the image which is highly sensitive to human eye, Cb represents the chroma blue component of the image and Cr represents the chroma red components of the image. These chroma components are not very sensitive to human eye. The conversion of RGB to YCbCr is illustrated as,

\[ Y = 16 + \frac{65.738R}{256} + \frac{129.057G}{256} + \frac{25.064B}{256} \]  
\[ Cb = 128 - \frac{37.945R}{256} - \frac{74.494G}{256} + \frac{112.439B}{256} \]  
\[ Cr = 128 + \frac{112.439R}{256} - \frac{94.154G}{256} - \frac{18.285B}{256} \]

Where R – Strength of red light,
G – Strength of green light,
B – Strength of blue light.

CIELUV colour space implements perceptual uniformity of the colours in the colour space. The CIELUV colour space represented the colours as combination of dimension rather than mixture of colours. The ‘L’ component represents the lightness or luminance of the colour, ‘U’ component and ‘V’ component represents co-ordinate measure position green/red and blue/yellow colour axes. The conversion of RGB to CIELUV is represented as,
Where $R$ – Strength of red light,
$G$ – Strength of green light,
$B$ – Strength of blue light,
$X, Y, Z$ – points in Cartesian coordinate system.

This paper utilizes VGG-Face CNN (Convolutional Neural Network) for feature extraction from colour space converted image. VGG-Face is a modified implementation of VGG-16, trained especially for face feature extraction. In this paper, VGG-Face’s classification layer is replaced by one global average pooling layer for feature extraction thus restricting to 19 layers deep CNN architecture with 512-length feature vector as output. The respective architecture is shown in Figure 4.

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = \begin{bmatrix}
0.412453 & 0.357580 & 0.180423 \\
0.212671 & 0.715160 & 0.072193 \\
0.019334 & 0.119193 & 0.950227
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\[L = \begin{cases} 
-116 \cdot Y^\frac{1}{2} - 16 & \text{for } Y > 0.008856 \\
903.3Y & \text{for } Y \leq 0.008856
\end{cases}
\]

\[u' = 4 \cdot \frac{X}{(X + 15Y + 3Z)}
\]

\[v' = 9 \cdot \frac{Y}{X + 15Y + 3Z}
\]

\[u = 13 \cdot L \cdot (u' - u_n) \text{ where } u_n = 0.19793943
\]

\[u = 13 \cdot L \cdot (u' - u_n) \text{ where } u_n = 0.19793943
\]

**Figure 4.** VGG-Face Architecture [11]

The face features extracted from VGG-Face is passed through Support Vector Classifier (SVC) for classification of real and spoof images. The detected face in YCbCr and CIELUV colour space is shown in Figure 5.
5. Experimental Setup

Face feature extraction and training and testing Support Vector Classifier (SVC) is done with a Quad-core Intel i5-8265U processor of 8GB RAM with base speed of 1.6 GHz and maximum speed of 3.9 GHz with Smart Cache of 6 MB. Deep learning toolkit namely tensorflow and keras and python modules namely open-cv, scikit-learn, mtcnn and vggface are used.

6. Results

Evaluation of the proposed method for spoof detection using various performance metrics is discussed in this section. Performance metrics include Accuracy, Specificity, Sensitivity and AUC-ROC Curve. Accuracy measures the total number of correct predictions. Specificity and Sensitivity computes the overall number of correct negative and positive predictions respectively. The ROC (Receiver Operating Characteristics) Curve represents the correlation between TPR vs FPR which differentiate the classifier ability at different thresholds as well as compared with No-skill classifier. The AUC (Area Under Curve)-ROC represents the likelihood of an input being correctly classified and it varies between 0-1.

| CLASS  | TRAIN | TEST |
|--------|-------|------|
| REAL   | 2777  | 1924 |
| SPOOF  | 2904  | 1864 |
| TOTAL  | 5681  | 3788 |

The face features are extracted from the YCbCr and CIELUV converted image using VGG-Face with pre-trained weights. These extracted features act as input for SVC (Support Vector Classifier). Training and testing images for SVC with respect to each class is illustrated in Table 2. Fine-tuned SVC is evaluated on a totally unseen data to generalize how well the classifier is performing. The confusion matrix shows that there were 14 misclassified images in Figure 6. Evaluation metrics based on the predictions in test data is illustrated in Table 3.
From the Figure 7 depicting the ROC Curve, it can be inferred that the classifier has good probability of classifying the input correctly since the area under curve is 0.9996. The dashed blue line indicates the classifier with No-skill. In other words, it is a classifier that always predicts 0.

### Table 3. Evaluation Metrics

| Class  | Precision | Sensitivity | Specificity | f1_score |
|--------|-----------|-------------|-------------|----------|
| Real   | 0.995     | 0.996       | 0.995       | 0.995    |
| Spoof  | 0.996     | 0.995       | 0.996       | 0.995    |

**Time:** 0.16 seconds

**Accuracy:** 99.6%

**False Acceptance Rate (FAR):** 0.41%

**False Rejection Rate (FRR):** 0.32%

**Half Total Error Rate (HTER):** 0.368%

**Equal Error Rate (EER):** 0.369%

Figure 8 demonstrates the output of spoof detection of single face. For multiple face spoof detection, the process is repeated for every detected face in the image. Multiple face spoof detection is shown in Figure 9. Comparison of results with the other works is illustrated in Table 4.
Figure 8. Single Face Spoof Detection

Figure 9. Multiple Face Spoof Detection
Table 4. Comparison of results

| Related Works | Dataset           | EER (%)  |
|---------------|-------------------|----------|
| Yousef et al. [1] | MSU-USSA         | 0.35±0.19|
| Lie et al. [2]      | CASIA             | 4.5      |
| Zinelabidine et al. [3] | Replay attack | 0.4%     |
| Jukka et al. [8]    | CASIA and NUAA    | 6.8      |
| **Proposed method** | **Custom and NUAA** | **0.369%** |

7. Conclusion
The Support Vector Classifier (SVC) have been able to classify real and spoof faces with an accuracy of 99.6%, FAR of 0.41% and FRR of 0.32% with face features extracted from YCbCr and CIELUV colour space converted images. The inadequate illumination affects the level of accuracy of the classification when it comes to real time application. The VGG-Face model used in extracting the face features was capable of apprehending the disparity among spoof and real faces in YCbCr and CIELUV colour spaces. It improves the security measures of the authentication with low cost and high accuracy and without the need of any additional hardware component.

8. References
[1] Atoum Y, Liu Y, Jourabloo A and Liu X 2017 Face anti-spoofing using patch and depth-based CNNs *IEEE International Joint Conference*
[2] Lie L, Feng X, Boulkanafer Z, Xia Z, Li M and Hadid A 2016 An original face anti-spoofing approach using partial convolutional neural network 6th *International conference on image processing theory*
[3] Boulkanafer Z, Komulainen J and Hadid A 2015 Face anti-spoofing based on colour texture analysis *IEEE International Conference on Image Processing (ICIP)*
[4] Balaji S, Guruprasad S and Sathiesh Kumar V 2019 Face-Spoof detection system using convolutional neural network *International Conference on Recent trends in Electronics, Computing and Communication Engineering (ICRTECC)*
[5] Määtta J, Hadid A and Pietikäinen M 2011 Face spoofing detection from single images using micro-texture analysis *International Joint Conference on Biometrics (IJCB)*
[6] Killiougli M, Taskiran M and Kahraman N 2017 Anti-spoofing face recognition with liveness detection using pupil tracking 15th *International Symposium on Applied Machine Intelligence and Informatics (SAMII)*
[7] Kant C and Sharma N 2013 Fake face recognition using fusion of thermal image and skin elasticity *International Journal for Computer Science and Information Security*
[8] Komulainen J, Hadid A and Pietikainen M 2013 Context based face anti-spoofing 6th *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*
[9] Zhang K, Zhang Z, Li Z and Qiao Joint Y 2016 Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks *IEEE Signal Processing Letters*
[10] Parkhi O M, Vedaldi A and Zisserman A 2015 Deep Face Recognition *British Machine Vision Conference*
[11] Nguyen V and Bai L 2010 Cosine similarity metric learning for face verification 10th *Asian conference on computer vision* (Queensland, New Zealand)