ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Face Spoofing Detection Using HIK Method of Codebook Generation

Christeena Mathews[1], Ambily Balaram[2]

M. Tech Student, Department of Computer Science and Engineering, KMCT College of Engineering, Kerala, India[1],

Assistant Professor, Department of Computer Science and Engineering, KMCT College of Engineering, Kerala, India[2]

ABSTRACT: Face recognition is being used in a variety of applications, but it's commonly used for authentication purpose. Even though it seem to be a very secure method, there are actually many loopholes. A face recognition system can be easily deceived by presenting a photograph or a digital video clip or a 3D mask in front of the camera, which will cause the system to recognize the face in the photograph or video as the face of a valid user. Such unauthorized intrusions are known as face spoofing attacks. In this paper we have proposed a method that uses the noise signature imprinted in the recaptured video for detecting such face spoofing attacks. A recaptured image will have noises added to it like blurring effects, printing artefacts, banding effects, Moir´e patterns etc. So we are extracting noise from the input video and analysing the temporal and spectral information present in it. For this process we have a two-tier feature extraction which include the low-level and mid-level features. For mid-level feature extraction a visual codebook is generated. Common practice for codebook generation is by k-means method, but in this paper we have used the Histogram Intersection Kernel (HIK) method instead of Euclidean distance in k-means. This method gives 2-4% more accuracy than standard k-means.

KEYWORDS: Spatio-temporal methods, residual noise, Fourier transform, feature extraction, recapturing

I. INTRODUCTION

Information security has become a high prioritized concern of every individual surfing on internet. This is because data theft is increasing day by day. Different modes of security steps are taken to avoid this stealing. Many user authentication techniques are being used. Knowledge-based and token-based methods are most widely used to date. But these methods use the information they possess or know for authentication and not who they are. So the stealing of such data can make the system vulnerable. Biometrics is an alternate authentication method which uses the behaviour of humans like fingerprint, face, iris, DNA, hand geometry etc. Face recognition method is the most significant of all the biometric methods available with different approaches like texture-based, shape-based, motion-based etc. The process of deceiving face authentication methods by presenting a fake sample like photograph or a video or a 3D model of an authorised person in front of the camera to obtain access is known as spoofing attack. Even though the technologies have advanced so much, the vulnerability of face recognition system against spoofing attack is still a serious issue.

The input captured by the camera in the face recognition system will definitely have some noise engraved into it. From various studies we know that, photograph or a video played on a digital device will cause a noise to get imprinted into it while recapturing. The noise is imprinted in the form of blurring effects, banding effects, Moire patterns etc. A spatio-temporal method is introduced in this paper that extracts such effects to provide an efficient way to differentiate between valid access and spoofing attempts. This is a low cost method that is independent of any user interaction or extra hardware. For this face spoofing detection method, time-spectral descriptors are extracted which include the temporal and spectral information from face biometric inputs.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

II. RELATED WORK

The existing face spoofing detection techniques can be divided into four types – user behaviour modelling, user cooperation based, methods that depend on additional hardware and methods that are based on data driven characterisation. The first one captures the behaviours of the user like eye-blink or slight head and face movements. This method works well against a photograph based or 3D mask based attack, but fails if the attacker is playing a digital video of the valid user. User cooperation based methods can be used to detect face spoofing by asking some challenge questions or by asking the user to perform specific movements. Specific movements asked to perform may be a facial expression, mouth movement or head rotation. Then the user's response is analysed to check whether the requested kind of movement is what the user actually performed. This kind of method is used because humans tend to be interactive, but a photograph or a video doesn't react. Hardware based solutions are very effective because they provide additional information on the surface reflectance and the depth of the observed sample. The disadvantage with hardware is the difficulty in implementation in mobile phones and tablets as hardware may not be supported. However hardware based methods may be kept under consideration because sensors may be emerging in mobile phones as well, in the near future. Finally data driven characterization based methods make use of only the data captured from the sensors that may give any hint on attempted attacks.

A. Frequency based approaches

Pinto [2] proposed a method for detecting video based spoof attacks using visual rhythm analysis. It was observed that in a video-based spoofing attack, a noise signature is added to the original sample while recapturing the videos. So the authors separated the noise using a low-pass filter and captured the temporal information using visual rhythm analysis.

Lee [3] proposed a method which was based on the frequency entropy of image sequences. The authors used face verification algorithm to recognize the face region. Then the RGB channels are normalized using z-score technique and then the independent component analysis method is applied to remove the cross-channel noise caused by interference from the environment. Finally to analyse the entropy of channels individually, authors calculated the power spectrum.

Pinto [4] also proposed a method by extracting time-spectral feature descriptors that gathers temporal andspectral information across the biometric sample and use thevisual codebook concept to find mid-level feature descriptorscomputed from the low-level ones.

B. Texture based approaches

A solution was proposed to handle attacks with printed photographs based on the change in surface roughness of an attempted attack and real face. The luminance and reflectance of image was estimated and classified those using Sparse Low Rank Bilinear Logistic Regression methods. Another area of analysis was the micro textures by using Local Binary Pattern (LBP). A method was proposed to extract different information from face like colour, texture and shape which will give a holistic representation of face. The variations in LBP operators were investigated and histograms were generated. To deal with mask based attacks solutions based on reflectance was considered. For this analysis, the decomposition of images into illumination and reflectance was done using Variational Retinex algorithm.

C. Motion based approaches

The motion information are extracted from the scene using two types of analysis – static and video-based analysis. Static analysis is by combining different visual features like colour, edge, and Gabor textures. Whereas video-based analysis uses simple motion related measures such as eye-blink, mouth movement and change in facial expression.

III. PROPOSED METHOD

A face spoofing detection method based on noise is introduced in this section. There are three main steps in this method: low-level descriptor extraction, mid-level descriptor extraction and classification. Fig 1 shows the main contents of these steps; they are explained in detail in the following sections.

This method is designed based on the fact that a noise pattern different from those found in real biometric samples is generated during the manufacture and recapture of the synthetic biometric samples. Määttä [5] and Tan [6] have found that there is a loss of some high frequency components while generating photographs for spoofing attacks. Pinto

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

[2] has reported in his works that due to the blurring patterns added while recapturing the sample displayed in tablets or smartphones or laptop screens, there is a significant increase in low frequency components. Apart from blurring effect, there are other artefacts added as noise like banding effect, flickering effect and Moiré pattern [7].

The proposed solution takes advantage of noise added to the biometric sample called noise signature. A Fourier analysis is performed on the noise signature to get information in the form of frequency, amplitude and phase of the component sinusoids. In this solution Fourier spectrum will quantify the flickering effect, blurring effect and Moire pattern. This system uses a two-tier feature extraction – low and mid-level. The noise signature is first transformed into frequency representation and time spectral features are extracted to form low-level descriptors. The concept of visual codebooks is used for converting the feature values into mid-level representation. The novelty of this method lies in the HIK based method for distance calculation in codebook generation. Instead of Euclidean distance which is usually used, Histogram intersection is used which gives 2-4% more accuracy.

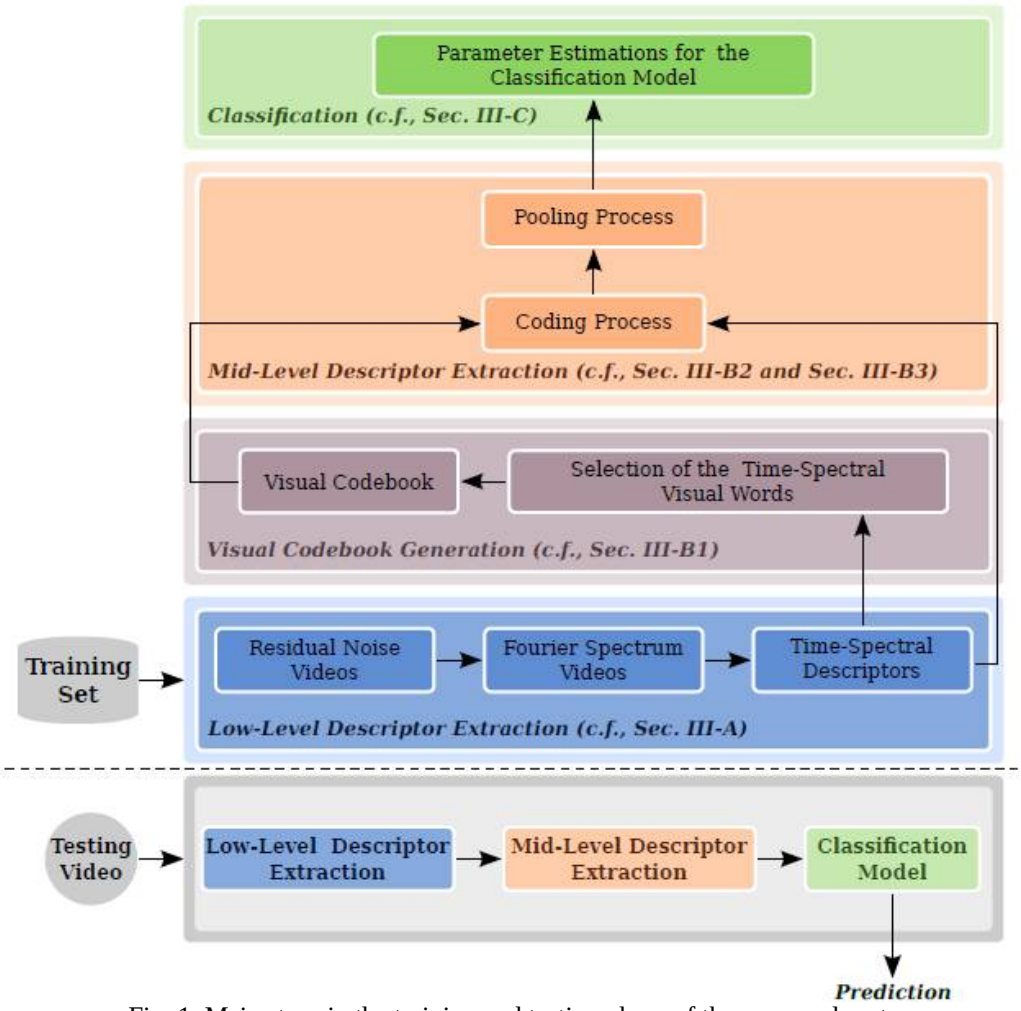


Fig. 1. Main steps in the training and testing phase of the proposed system.

A. Low-level Descriptor Extraction

Pinto et al [1] have found from his works that noise signature gives significant information for spoofing detection. The steps used in low-level descriptor extraction are:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

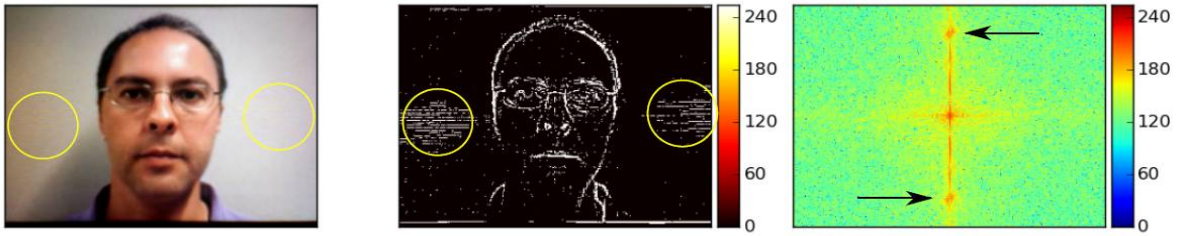
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

1. *Calculation of the Residual Noise Videos:* A Gaussian filter is used for extracting noise from a given video by removing the high frequency components. Then the filtered video is subtracted from the input video generating a Residual Noise Voice (V_{RN}).
2. *Calculation of the Fourier Spectrum Videos:* From the residual noise video, possible artifacts and noise patterns can be analyzed by applying the 2D Discrete Fourier Transform to each frame of the V_{RN}. In this method, the magnitude spectrum and phase spectrum of the noise signal is evaluated. The result is a Fourier Spectrum Video. Spectral and temporal information can be extracted from the Fourier Spectrum video by capturing the peaks present in the central region which are caused by blurring effect, banding effect and Moiré pattern.
3. *Computation of time-spectral descriptor:* For feature descriptor calculation, n temporal cubes are gathered from an input video. Computation of measure can be either on each frame or between consecutive frames. At the end of this process, a set of n time-spectral descriptors of (t-1) dimensions is generated for each video when spatio-temporal measures are applied.

a. Valid access video, its residual noise video, and corresponding magnitude spectrum.



b. Mobile attack video,its residual noise video, and corresponding magnitude spectrum.

c. Another mobile attack video,its residual noise video, and corresponding magnitude spectrum.

Fig 2. Examples of valid access and spoof attempt video frames, their corresponding residual noise video and magnitude spectra.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Examples of valid access video frame and mobile attack video frame, their residual noise video frames and their corresponding magnitude spectrum are shown in Fig 2.a, 2.b, and 2.c.

B. Mid-level Descriptor Extraction

In order to convert low-level features into mid-level the concept of Bag-Of-Visual-Words model is used. There are three steps in the process of extracting mid-level features: visual codebook generation, coding, pooling.

1. ***Visual Codebook generation:*** The generation of the visual codebook consists in the selection of time-spectral descriptors that are more frequent and representative considering all descriptors extracted from training videos. Theselected descriptors, called time-spectral visual words, form the visual codebook. The selection can be performed using two strategies: (1) random selection, whereby all descriptors are pooled and m visual words are randomly chosen using a uniform distribution; or (2) selection via clustering (e.g., k-means) whereby all descriptors undergo a clustering process and the m centroids found by the algorithm are used to form the visual codebook. In both cases, we end up with a single visual codebook, which is used to encode the low-level time-spectral descriptors from videos.

   In this paper we use Histogram Intersection Kernel (HIK) instead of Euclidean distance in k-means algorithm for codebook generation. The HIK codebook has consistently higher recognition accuracy over k-means codebook by 2-4%.

   Instead of pooling all descriptors extracted from videosinto a training set to build a single visual codebook, we canbuild class-based visual codebooks. When creating class-basedvisual codebooks, we consider the use of valid access and attempted attack video descriptors separately in order to findcodebooks in each class. For each class-based codebook, weuse the same procedures described above for a single visualcodebook creation. The two visual codebooks are concatenatedto create the final codebook.

2. ***Coding:*** The coding process performs a point wise transformation of the low-level descriptors into another representation. There are several strategies for coding being the hard and soft assignments the most common. Given a visual codebook and a low-level descriptor, the hard assignment transforms such descriptor into a binary vector with only one nonzero coefficient representing the visual word closest to it. The soft assignment, in turn, gives a real valued vector that represents the descriptor as a linear combination of the visual words of the codebook, whose coefficients give an associativity degree between the descriptor and the visual words of the codebook. In this paper, we evaluate thesetwo strategies for coding the low-level descriptors.

3. ***Pooling:*** The pooling process aims at summarizing the information contained in the set of n mid-level feature descriptors extracted from an input video into only one feature descriptor to obtain its final representation. In the literature, we have two common techniques to do that, known as sum-pooling and max-pooling. In this paper, we use max-pooling.

C. Classification:

After finding a new space representation for the videos in the database, we use machine learning algorithms to find a classification model to decide whether a sample is an attempted attack or a valid access. In this paper we use Support Vector Machine (SVM) algorithms.

IV. EXPERIMENTAL RESULTS

The UVAD dataset is considered for training in this paper. This dataset contains valid access and attempted attack videos of 404 different people, all created at full HD quality, 30fps, and nine seconds long. It contains 16,268 attempted attack videos and 808 valid access videos. Fig 3 and 4 illustrates some examples of this dataset.

In the low-level descriptor extraction phase noise signature from RGB videos are extracted using a Gaussian filter. Next, we extract cuboids of size 32x32x8 from the Fourier spectrum videos. The use of spatio-temporal measures produces 21 dimensional low-level feature descriptor. The proposed method presents better results using time-spectral

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

features extracted from magnitude spectrum videos considering the whole frames of a video and using the correlation measure for generating time-spectral descriptors.

The class-based codebooks outperform the single codebook and the best selection strategy is k-means clustering. The most appropriate size for codebooks is 320 visual words and the softmax outperformed other coding and pooling processes. Instead of Euclidean distance based k-means algorithm for codebook generation, in this paper we are using HIK based method which has a 2-4% higher accuracy than k-means and has almost the same complexity.

With this configuration, we obtained an AUC of 76.1125% and an HTER of 26.04%. The proposed method outperforms the ones based on texture analysis and also the methods based on motion analysis. The proposed method takes about 2 seconds to detect an attempted attack.

Fig 3. Examples of valid access video frames from UVAD dataset

Fig 4. Examples of attempted attack video frames from UVAD dataset.

V. CONCLUSION AND FUTURE WORK

In this paper we proposed a face spoofing detection method that makes use of the noise that gets added to the biometric sample while recapturing. The novelty of the paper lies in the concept of HIK based distance calculation in k-means for codebook generation. The experimental results showed that the HIK based method performs better than the other existing methods. The proposed method gave 2-4 % more accuracy against the state-of-the-art methods, whereas it has the same complexity of standard k-means method. The accuracy reflects in the HTER calculated and the comparison with the existing methods is depicted in Table I. However this solution may not be the best anti-spoofing method, but it's a significant advancement in spoofing detection mission. More research and experiments on it by combining with the existing methods will make it the best solution.

TABLE I
Comparison among the proposed method and different existing methods

Methods	FAR (%)	FRR (%)	HTER (%)
LBP based	27.41	66.04	46.72
Correlation based	81.6	14.56	48.06
Using time-spectral cubes	28.29	28.33	28.31
Proposed method	**24.13**	**27.95**	**26.04**

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

REFERENCES

1. A. Pinto, H. Pedrini, W. R. Schwartz and A. Rocha, "Face Spoofing Detection Through Visual Codebooks of Spectral Temporal Cubes," in *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726-4740, Dec. 2015.
2. A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Video-based face spoofing detection through visual rhythm analysis," in Conference on Graphics, Patterns and Images, Aug. 2012, pp. 221–228.
3. T.-W. Lee, G.-H. Ju, H.-S. Liu, and Y.-S. Wu, "Liveness detection using frequency entropy of image sequences," in IEEE Int. Conference on Acoustics, Speech, and Signal Processing, 2013, pp. 2367–2370.
4. A. Pinto, H. Pedrini, W. R. Schwartz and A. Rocha, "Face Spoofing Detection Through Visual Codebooks of Spectral Temporal Cubes," in *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726-4740, Dec. 2015.
5. J. M¨a¨att¨a, A. Hadid, and M. Pietik¨ainen, "Face spoofing detection from single images using micro-texture analysis," in IEEE Int. Joint Conference on Biometrics, Oct. 2011, pp. 1 –7.
6. X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in European Conference on Computer Vision, 2010, pp. 504–517.
7. A. Beach, Real world video compression, 1st ed. Berkeley, CA, USA: Peachpit Press, 2008
8. C. Xu, Y. Zheng, and Z. Wang, "Eye states detection by boosting local binary pattern histogram features," in IEEE Int. Conference in Image Processing, Oct. 2008, pp. 1480–1483.
9. I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in Int. Conference of the Biometrics Special Interest Group, Sep. 2012, pp. 1 –7.
10. N. Kose and J.-L. Dugelay, "Reflectance analysis based countermeasure technique to detect face mask attacks," in Int. Conference on Digital Signal Processing, 2013, pp. 1–6.
11. N. Erdogmus and S. Marcel, "Spoofing 2D face recognition systems with 3D masks," in Int. Conference of the Biometrics Special Interest Group, 2013, pp. 1–8.
12. A. Pinto, "A countermeasure method for video-based face spoofing attacks," Master's thesis, University of Campinas, Oct. 2013.

BIOGRAPHY

Christeena Mathews received her B.Tech degree in Information Technology from Cochin University of Science and Technology, Cochin, in 2012 and doing M.Tech in Computer Science at KMCT College of Engineering, Calicut. Her areas of research interest include image processing and security.

Ambily Balaram received her B.Tech degree in Computer Science and Engineering from Kannur University, Kerala in 2007 and her M.Tech degree in Computer Science and Engineering from Calicut University in 2013. She has more than 6 years of experience in teaching field and is currently working as Assistant Professor in KMCT College of Engineering, Calicut. Her research area is data mining.