Application of Spectral Information in Identification of Real-Fake Face Images

Akhilesh Kumar Pandey
Natinal Institute of Technology, Kurukshetra,
Haryana-136119
er.akhileshpandey@yahoo.com

Rajoo Pandey
Natinal Institute of Technology, Kurukshetra,
Haryana-136119
rajoo_pandey@nitkkr.ac.in

Abstract

The biometric authentication systems based on face recognition are easy to implement with any device, which has an in-built camera. These systems are very secure, but with various attacking methods, such security also becomes vulnerable. Among various types of possible attacks, spoofing is an attack in which available face information is presented before the sensor to mislead the authentication system. In this paper, a model has been presented based on spectral analysis of the captured images to classify them as real faces and face images. We consider Fourier and cosine transform of the image for evaluation of various Image Quality Measures (IQMs), with which the classification is performed using neural networks. The proposed spectral contents based IQMs are compared with several conventional IQMs to judge their performance. The simulation on Replay-Attack database has shown the improvement in the performance of the present model.

General Terms PSNR, SNR, SME, SPE.

Keywords Image Quality Measures, Real-Fake Discrimination, Neural Networks.

1. Introduction

There has been a remarkable change in the authentication systems in recent years with application of biometrics replacing the conventional systems. The biometric systems based on fingerprint is one of the most widely used authentication mechanism in this decade. However, the implementation of such systems is costlier if used in laptops/tabs because of the requirement of an additional sensor and its supportive environment. The hand-held consumer devices generally have built-in cameras, hence the face recognition based authentication mechanism can be easily implemented in such appliances, without changing their cost significantly. But, one of the major drawbacks of using such authentication mechanism is the possibility of spoofing attack. Although, face recognition is very popular due to its non-intrusive nature, its disadvantage lies in its that very merit, as any eavesdropper can capture the image of a person (without letting him know) and can present it in front of the camera for authentication [2, 7, 9].

The main focus of the present work, is to maximize the discrimination between real and fake sample to avoid spoofing in face recognition based authentication system. In this paper, we attempt to explore Fourier and cosine transform characteristics of the captured image and NN classification for better recognition rate.

This paper is organized as follows. In section 2, a brief review of the literature is presented. In section 3, proposed architecture of the model has been presented. and experimental results are shown in section 4. The conclusion has been drawn in section 5.

2. Related Work

In order to detect the liveness of the face, one of the frequently used methods is to employ either multiple cameras to capture 3-D face image or multiple stage authentication. However, these mechanisms guarantee the authentication, but tend to make systems bulky and costly. Also, it can be perceived that the multiple sensor based method becomes more sophisticated, while implementing in hand-held devices.

As spoofing detection is challenging, it has gained attention of several research communities. Most of the techniques are based on estimation of motion of the face, and are therefore well suited for detection of print type spoofing attacks. Also, scenic clues of faces and relative motion of the gestures are widely studied for making discrimination between real face and face image. Such approaches become vulnerable under the video type attack, in which a video is presented in front of the camera for authentication. Various other methods such as analysis of the motion under Fourier spectra, instantaneous quality assessment of the captured image are useful for real fake discrimination. Based on the above mentioned information, we classify the anti-spoofing technique in two categories: *(i)* Motion-based techniques, which involves directly or indirectly the computation of motion and, *(ii)* Other methods, which do not require any information regarding the motion.

1. In order to detect the liveness of the face, the motion of the eye-blinks has been studied in [10, 11], which requires user cooperation and hence results in to a time taking process. In [8], first, the motion is estimated as number of frames increases, then Fourier transform is generated with this series. After this different parameters from the Fourier spectra are computed to classify the incoming video from camera as real or fake. It is obvious that the natural motion of the human face captured by a fixed camera would be different than in case of a hand-held camera. It is expected to be motionless in the frames whereas in the later case it is full of motion.
2. In [14], the Lambertian model is used to study the surface properties of one sample, which is further followed by sparse

ICCCT '15,, September 25-27, 2015, Allahabad, India.
Copyright © 2015 ACM 978-1-4503-3552-2/15/09. . . $15.00.
http://dx.doi.org/10.1145/2818567.2818606 Reprinted from ICCCT '15,, [Unknown Proceedings], September 25-27, 2015, Allahabad, India, pp. 1–4.

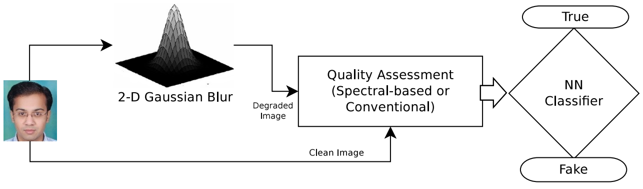


Figure 1. The system model used used here

logistic regression to make a binary decision about the liveness of the face. It has been shown in studies that quality of the original image will differ from its copy [13]. Therefore, image quality assessment is done for classification of the captured face image, as real or fake [5]. The authors have evaluated 25 different quality measures such as peak signal to noise ratio (PSNR), structural content (SC) etc, using a Gaussian blur, and further employed LDA for classification.

3. System Model

The basic system model is shown in figure 1. The system model has two basic stages; the first stage employs the degradation filter to compute the distorted image and the second stage uses the transform in the original and degraded image to extract various features described later in this section. Using these features, the classification is performed with the help of a multi-layer perceptron (MLP).

3.1 Degradation Model

The average filter smooths the image without preserving the details and without considering the characteristic of the neighborhood pixels in an image. Therefore, the Gaussian filters have wide range of applications in various image processing fields. A two-dimensional Gaussian filter is described as:

$$h(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{1}$$

where, $h(x,y)$ is normalized filter coefficient at location (x,y), and σ is standard deviation of the filter.

The filter used in [5], is of the size of 3×3, having $\sigma = 0.5$. In this paper, the original image captured from the camera is represented as I and the degraded image is represented as $\hat{I} = I * h$.

3.2 Transforms Definitions

Several transforms have been used in image processing for various applications. Here, we consider Fourier and cosine transforms for the analysis of the real and fake images. The Fourier transform has wide range of application in fields like denoising, image reconstruction, pattern recognition etc, whereas the cosine transform does not involve complex computation, and is used in image compression. The 2-D Fourier transform and cosine transform are described, respectively, as:

$$F(k,l)=\sum_{m=0}^{m=M-1}\sum_{n=0}^{n=N-1}\left\{I(m,n)\exp\left(-j2\pi\left(\tfrac{km}{M}+\tfrac{ln}{N}\right)\right)\right\} \tag{2}$$

and

$$C(k,l)=\sum_{m=0}^{m=M-1}\sum_{n=0}^{n=N-1}\left\{I(m,n)\cos\left(\tfrac{(2m+1)k\pi}{2M}\right)\cos\left(\tfrac{(2n+1)l\pi}{2N}\right)\right\} \tag{3}$$

Where, F and C are Fourier and cosine transform of the image I, respectively, having size of $M \times N$.

3.3 Conventional quality measures

Several quality measures are available in the literature which are used in various applications like contrast enhancement, image inpainting, reversible data hiding etc. Few commonly used quality measures are Mean Square Error (MSE), Peak Signal to Noise ratio (PSNR), Signal to Noise Ratio (SNR), Normalized Absolute Error (NAE) and Structural Content (SC), which are described, respectively, as:

$$MSE\left(I,\hat{I}\right)=\tfrac{1}{MN}\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(I_{i,j}-\hat{I}_{i,j}\right)^2 \tag{4}$$

$$PSNR\left(I,\hat{I}\right)=10\log\frac{\max\left(I^2\right)}{MSE\left(I,\hat{I}\right)} \tag{5}$$

$$SNR\left(I,\hat{I}\right)=10\log\frac{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(I_{i,j}\right)^2}{M*N*MSE\left(I,\hat{I}\right)} \tag{6}$$

$$NAE\left(I,\hat{I}\right)=\frac{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left|I_{i,j}-\hat{I}_{i,j}\right|}{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(I_{i,j}\right)} \tag{7}$$

and

$$SC\left(I,\hat{I}\right)=\frac{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(I_{i,j}\right)^2}{\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(\hat{I}_{i,j}\right)^2} \tag{8}$$

More details can be found in [1, 4, 6, 15].

3.4 Quality assessment through the transforms

The quality of the image can be assessed with the transforms and also without them. Here, different quality measures are considered using the transforms. The Spectral Magnitude Error (SME), Spectral Phase Error (SPE) and High Low Frequency Index (HLFI) are based on the Fourier transform and described, respectively, as:

$$SME\left(I,\hat{I}\right)=\tfrac{1}{MN}\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(|F_{i,j}|-\left|\hat{F}_{i,j}\right|\right)^2, \tag{9}$$

$$SPE\left(I,\hat{I}\right)=\tfrac{1}{MN}\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(\left|\arg(F_{i,j})-\arg\left(\hat{F}_{i,j}\right)\right|\right)^2, \tag{10}$$

and

$$HLFI(I)=\frac{\sum_{i=1}^{i_l}\sum_{j=1}^{j_l}|F_{i,j}|-\sum_{i=i_l+1}^{M}\sum_{j=j_l+1}^{N}|F_{i,j}|}{\sum_{i=1}^{M}\sum_{j=1}^{N}|F_{i,j}|} \tag{11}$$

Where, F and $\hat{F}$ are the Fourier transform of I and $\hat{I}$, respectively. $i_l = 0.15M$ and $j_l = 0.15N$ are lower frequency thresholds. More details can be found in [5].

In the cosine domain, we propose two new quality measure has been introduced for real fake discrimination. The Difference in Cosine Domain (DCD) and Loss in Cosine Domain (LCD) are described, respectively, as:

$$DCD\left(I,\hat{I}\right)=\tfrac{1}{MN}\sum_{i=1}^{i=M}\sum_{j=1}^{j=N}\left(|C_{i,j}|-\left|\hat{C}_{i,j}\right|\right)^2 \tag{12}$$

and

$$LCD(I)=\tfrac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{i,j}-\widetilde{I}_{i,j}\right)^2 \tag{13}$$

Where, C and $\hat{C}$ are the cosine transform of the image I and $\hat{I}$, and $\widetilde{I}$ is computed as the inverse cosine transform of the C which is thresholded for the compression as $C_{i,j} = 0$, $\forall\left|C_{i,j}\right| \leq 0.5$.

3.5 Use of multiple frames

In the present study, quality of multiple frames is also assessed for improving the real fake discrimination. Previously described quality measures are computed for every frame and then liveness decision is made. All the outcomes are combined to make the final decision about their liveness based on the majority of all outcomes. The increase in number of frames will produce the higher recognition rate. This can be seen easily through the binomial distribution.

If we have total $N = 2k+1$ frames, then probability of getting at least $k+1$ identical decisions is given by:

$$P(X \geq k) = \Sigma_{x=k}^{N} \left(\begin{array}{c} N \\ x \end{array} \right) (p)^{x} (q)^{N-x} \quad (14)$$

where, k is a positive integer, p is probability of true recognition and $q = 1 - p$, N denotes the total number of frames, and $\left(\begin{array}{c} N \\ x \end{array} \right) = \frac{N!}{x!(N-x)!}$ is total possibilities of choosing x out of N.

3.6 Classification Strategy

We consider neural network based classification for real-fake identification. Neural Network have been widely used in several applications involving pattern recognition [12]. The NN considered in the present model has a hidden layer, having 5 nodes, with 5 inputs and one output. The hidden layer has nonlinear nodes, and employs tansigmoid function in the nodes, as defined by:

$$\nabla = \tanh(S) = \frac{e^{+S} - e^{-S}}{e^{+S} + e^{-S}} \quad (15)$$

Where, S denotes the activation sum for the nodes.

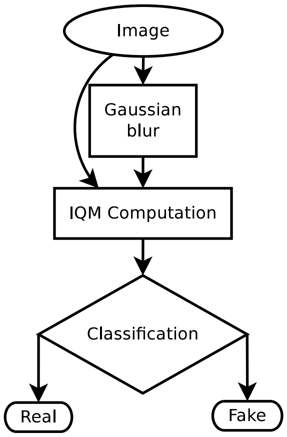
There are several NN training schemes that can be used to train the network with given inputs and targets such as Back Propagation, Scaled Conjugate Gradient, Levenberg-Marquardt (LM) etc. We consider LM algorithm for training as it has faster rate of convergence.

4. Experiments and Results

We used the Replay-Attack Database for the simulation of the proposed model in which various challenges are incorporated such as different quality, illumination and motion of the fake images [3]. Different quality in the fake image is introduced using different capturing method involving print, mobile video, tab video of the person etc. Motion is introduced using different support types, hand and fixed, as it can be easily perceived that using fixed support type will produce no motion, while the hand type will result abnormal motion in comparison to the real ones.

This database is divided in three main categories; training, development and testing, named as *Train*, *Devel* and *Test*. The *Train* set has 60 real videos and 80 videos of the fake for each support type, whereas *Devel* and *Test* sets have 80 real videos and 200 fake videos for each support type. The location of the face is also given in the database. With the help of these locations, facial part of the frame is cropped and resized to 64x64 pixels for further processing. Basic flow of the algorithm is shown in Figure 2.

Figure 2. Basic flow of the process of identification of real and fake face samples



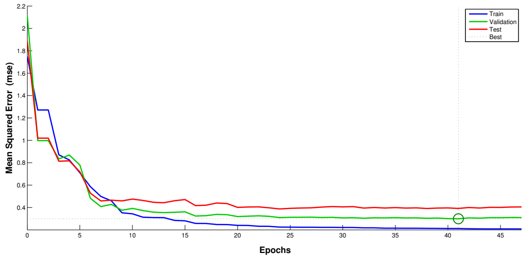


Figure 3. Mean Square Error vs number of epochs for the present Neural Network

The IQMs are then computed with Gaussian blur and performance is measured using NN classifier. The performance is observed in terms of False Acceptance Rate (FAR), False Rejection Rate (FRR) and Total Error Rate (TER). The FAR accounts for the fake samples that are accepted by the authentication system, while the FRR indicates for real samples that are rejected by the system.

In this model, we consider feed forward type neural network for the classification. In the experiment, 400 randomly selected samples are used to train the NN from the *Train* set of the database, in which 200 belong to real and 100 each belong to fake videos (for fixed and hand support types). The performance of the model is observed in all the sets of the database. The Convergence of the NN has been shown in figure 3. The trained NN is used for classification. The results obtained are shown in table 1, from which it can be observed that with the use of spectral content more discrimination can be achieved between real and fake samples.

Table 1. Various Errors on Different Test Sets using Conventional and Spectral IQMs

Test Sets	Conventional IQMs			Spectral IQMs		
	FAR	FRR	TER	FAR	FRR	TER
Train	9.04	5.81	14.85	7.38	3.11	10.49
Devel	11.56	8.56	20.12	7.85	3.74	11.59
Test	12.32	7.62	19.94	9.00	4.71	13.71

When we have image sequences and ample computational power, then the discrimination between the real and fake samples becomes easier. In the present work, maximum up to 9 frames are considered for the classification. These frames are individually analyzed to make the final decision about the liveness of the samples. Total five scenarios are considered in this work and results are shown in table 2. With this it can be seen that as the number of frames increases, the recognition accuracy also improves. It can also be observed that spectral information based IQMs offer higher discrimination than the conventional IQMs. For an error rate of 10% spectral information based IQMs require 3 frames in comparison to 9 frames in case of conventional IQMs.

Table 2. Total Error Rate on various IQMs under different number of frames on Test set of Replay-Attack database

Total Frames	Conventional IQMs	Spectral IQMs
1	19.94	13.71
3	15.45	10.18
5	13.21	8.47
7	12.16	7.31
9	10.86	5.92

5. Conclusions

In this paper, various IQMs are explored for making the discrimination between real and fake face images captured from the camera. The classification in this model is based on the neural networks. The proposed spectral information based quality measures provide higher discrimination than the conventional IQMs. In this study, it has been shown that spectral measures have approximately twice as much information about the discrimination in comparison to other IQMs. It has also been shown that if the enough computational power is available, then multiple frames can be used to detect the liveness of the sample more accurately. Use of the multiple frames enhances the recognition rate significantly. Other transforms can also be explored in future for further improvement in discrimination.

Bibliography

[1] I. Avcibas, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *Electronic Imaging*, 11(2):206–223, Apr. 2002.

[2] S. Bayram, İ. Avcıbaş, B. Sankur, and N. Memon. Image manipulation detection. *Journal of Electronic Imaging*, 15(4):041102–041102, 2006.

[3] I. Chingovska, A. Anjos, and S. Marcel. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. In *IEEE BIOSIG 2012*, Sept. 2012.

[4] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *Communications, IEEE Transactions on*, 43(12):2959–2965, 1995.

[5] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *Image Processing, IEEE Transactions on*, 23(2):710–724, 2014.

[6] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.

[7] A. K. Jain, K. Nandakumar, and A. Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing*, 2008:113, 2008.

[8] J. Li, Y. Wang, T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra. In *Defense and Security*, pages 296–303. International Society for Optics and Photonics, 2004.

[9] K. A. Nixon, V. Aimale, and R. K. Rowe. Spoof detection schemes. In *Handbook of biometrics*, pages 403–423. Springer, 2008.

[10] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[11] G. Pan, L. Sun, and Z. Wu. *Liveness detection for face recognition*. INTECH Open Access Publisher, 2008.

[12] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 38–44. IEEE, 1998.

[13] M. C. Stamm and K. R. Liu. Forensic detection of image manipulation using statistical intrinsic fingerprints. *Information Forensics and Security, IEEE Transactions on*, 5(3):492–506, 2010.

[14] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Computer Vision–ECCV 2010*, pages 504–517. Springer, 2010.

[15] S. Yao, W. Lin, E. Ong, and Z. Lu. Contrast signal-to-noise ratio for image quality assessment. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1. IEEE, 2005.