SECURITY ISSUES IN FACE RECOGNITION

Junied Khalid Khan
Amity School of Engineering
and Technology (ASET)
Amity University Noida, India
Junaid01khan@gmail.com

Divya Upadhyay
Amity School of Engineering
and Technology (ASET)
Amity University Noida, India
dupadhyay@amity.edu

***Abstract*- Data uncertainty in face recognition due to various environmental factors and spoofing face recognition with 3D- masks are two major threats to gain illegitimate access. A single image of face has got high uncertainty in representing the face since the face image varies with illumination, facial expression and pose. Thus a face image can only be considered as an observation but not an absolute accurate representation of the face. The accuracy of the face recognition can be improved by obtaining more face images from the same person. However in real-world we only have a limited number of available face images and thus there is high uncertainty. In this paper, we attempt to reduce the uncertainty of face image by synthesizing the virtual training samples. Similarly, spoofing is the act of gaining an illegitimate access by masquerading as a valid user by falsifying data. Among all biometric traits, face can be exposed to the most serious threats as it is easy to access and reproduce. In this paper various spoofing techniques have been have been examined and different algorithms have been put forward to detect them. We try to examine the spoofing potential of 3D facial masks for different recognition systems and address the detection problem for this complex attack.**

***Keywords*- Face recognition, Uncertainty, Spoofing, Mask attack, Presentation attack.**

I.INTRODUCTION

In real world applications data uncertainty is common due to various factors like inappropriate measurement and sampling errors [1]. There are various methods for processing uncertain data including uncertain data mining, uncertain data management [2] and uncertain data clustering [3]. The data uncertainty is represented in two main ways. The first representation of data is through probability distributions rather than deterministic values and the second approach is to represent data by statistical information like mean and variance.

Face recognition is most commonly used biometric trait by humans and has got number of applications in software and consumer electronics. Face can be easily accessed when compared with other biometric traits [4], such as fingerprint or iris. However in malicious circumstances this advantage becomes a drawback by enabling attackers to create copies and spoof into the face recognition system. Spoofing attack is the act of outwitting a face recognition system or any biometric system in order to gain authentication by presenting fake evidence [5]. Attacker can gain access by simply displaying the printed photos or replaying the recorded videos [6] of the valid user before the sensor.

Face recognition is one of the most attractive biometric techniques but still a challenging task. It has become challenging task due to the factors like varying light, environment, facial expression and pose which cause uncertainty in face recognition. As a result, one training sample cannot reflect all the variations thus it is just an observation. To attain high recognition accuracy, we can take number of training samples. However in real world applications, due to the limitation of storage space and image acquisition time only few training samples are available. Many popular face recognition applications like ID card identification, e-passport and law enhancement fail to work well in these conditions. To address this problem, we try to produce more training samples by synthesizing virtual training samples as more training sample means less uncertainty in face recognition system. Although we synthesize virtual training samples to reduce uncertainty of the face but at the same time it can lead to high computational cost which may contain lot of redundant information. [7]To address this problem sparse-representation based algorithm has been used to solve the uncertainty in face recognition. Sparse representation requires a test sample to be sparsely represented by a weighted sum of all the training samples.

The vulnerabilities like photo print and video replay attacks have evolved huge interest in the biometric community and many papers have been published to counter this problem. Existing anti-

978-1-4799-4236-7/14/$31.00 ©2014 IEEE

spoofing approaches for these type of attacks can be categorized into three types:

I. Texture analysis: Assuming the presence of blurring [8] many anti-spoofing techniques examine the texture of the captured face image. Similarly in a recent study another technique called micro-texture analysis [8] is proposed which uses multi-scale local binary patterns and it depends on the quality of the printed image and video.

II. Motion analysis: This method aims to detect spoofing attacks by analyzing the motion of the scene. [9]Since the planar objects like mobile phone or paper sheet move significantly in a different way as compared to real faces.

III. Liveness detection: Finally, liveness of the face is determined based on the gestures like lip movement or eye blinking. However these approaches are bound to fail with photographic masks being used which are high resolution facial prints worn after eyes and mouth regions are cutout. Such work on fraud detection is still limited and is based on the flatness of the captured surface in front of sensor during attack. [10]Now approaches where 3D nature of the face is examined by employing cameras like kinect are used. With the advancements in 3D technologies, facial masks are easily available to introduce the new paradigm in the spoofing attacks. To the best of our knowledge very few studies have been published so far to address this issue.

II.PROPOSED APPROACH

Sparse representation algorithms discussed earlier to solve data uncertainty in face recognition has got a serious drawback that it consumes huge amount of time. In contrast, l2-norm-based representation algorithm shows excellent results and is efficient in terms of recognition. The proposed approach consists of three steps. The first step puts forward the method of synthesizing virtual training samples which are useful in reducing the uncertainty of the face. [11]The second step proposes a scheme to determine from a set of all the original and synthesized virtual training samples the most useful training samples. The third step uses the l2-norm based representation algorithm to precisely classify the test sample.

In the first step, an effective way to reduce random uncertainty is to use limited training samples to synthesize virtual training samples which reflect wide variations of the face. Let us assume that X_i consists of n training samples and let $X_i = [x_i^1, \ldots x_i^n]$. Every two samples in these n samples are used to synthesize one virtual training sample and thus $c_n^2 = (n(n-1)) \square 2$ virtual training samples are produced for this ith class. If x_i^p and x_i^q are two original training samples from X_i, then corresponding virtual training samples are:

$$x_v^j = (x_i^p + x_i^q) / 2 \qquad \ldots\ldots (1)$$

Where j=1….. c_n^2. The c_n^2 virtual training samples of the ith class are $x_v^1 \ldots\ldots x_v^{c_n^2}$ respectively. The label of the original training samples of the ith class is taken as i. Blend training sample set (BTSS) is formed by combining all the original training samples and synthesized virtual training samples.

In the second step, we propose to select training samples that are similar to the test sample from BTSS. [12]These are the useful training samples which are used for the representation and classification of test sample. There are many measurements to select useful training samples from BTSS but for our convenience we take Euclidean distance as the measurement.

Let us assume that there are l training samples in BTSS from c classes and x^j (j=1….l; l>n) represent the j^{th} training sample in BTSS. Distance between y and x_j is calculated when a test sample y $\square$ R^m using:

$$dist_j = \| y - x_j \|_2 \qquad \ldots\ldots (2)$$

Eq. (2) gives the measurement of similarity between y and x^j. If the value of $dist_j$ is small, it means that x^j is similar to test sample y. First k smallest distances are identified as k useful training samples and rests of the samples are discarded. For k useful training samples obtained, a set of labels is defined as $S = \{s_1 \ldots s_d\}$, which is set of some numbers.

In the third step of proposed approach l2-norm based representation algorithm is represented as a linear combination of useful training samples k. Its aim is to minimize representation coefficients of l2-norm vector. Thus objective function can be written as:

$$\min_{ß`} \|y - x`ß`\|_2^2 + \theta\|ß`\|^2_2 \qquad \ldots. (3)$$

The selected k useful training samples are represented as $x` = [x`_1 \ldots x'_k]$ and the representation coefficients are given as $\beta` = [\beta`1 \ldots . \beta`k]$. θ is called regularization parameter and it is useful for the stability of the least square solution. This approach is also used to obtain the following equation:

$$` = (x`^T x` + I)^{-1} x`^T y \qquad \ldots\ldots (4)$$

Identity matrix is represented as I. The sum of the contribution to represent the test sample of the r^{th} class if all the useful training samples from the r^{th} class are $x_t^{-p} \ldots\ldots x_t^{-q}$ then

$g_r = \grave{}_p x_t^{-p} + \ldots + \grave{}_q x^{-q}$ … (5)

Where the coefficients of $x_t^{-p} \ldots\ldots x_t^{-q}$ are represented by $\grave{}_p \ldots\ldots \grave{}_q$ respectively. Residual of y_r from y is calculated by

$D_r = \|y - g_r\|^2$ ……(6)

The lesser value of D_r means greater contribution for the representation of test sample.

In case of spoofing, some earliest studies aim to distinguish between mask material and facial skin by detecting the difference in their reflectance.[13] But it is not valid now even plastic surgeries can be detected as it reduces the thermal signature of face. Later Kose et al published his studies for spoofing with masks for which a non-public database comprised of printed 3D masks of 16 users is utilized. The database called Morpho database constructs the face models of clients with 3D laser scanner and 3D printing is used to manufacture the masks. There are two spoofing databases for which 3D masks are used to generate the attacks.

I. Morpho Database: Morpho Database is a non-public database. It contains 207 real access and 199 mask attack samples in both 2D and 3D i.e. facial images and 3D-models. Masks used for spoofing are manufactured using facial models of 16 different users by a 3D printer. The shapes of masks are exact replicas of the targeted client.[14] A 3D scanner is used to acquire different shots for each user.

II. 3D Mask Attack Database: A face spoofing database called 3D Mask Attack Database (3DMAD) contains 76500 frames of 17 different users obtained through kinect sensor from Microsoft for both real access and attack based on spoofing using 3D facial masks.[15] Each frame contains following:
    a) A depth image of size ranging from 840 × 480 pixels to 1 × 11 bits.
    b) Corresponding color image of size ranging from 640 × 480 pixels to 3 × 8 bits.
    c) Manually ammotated eye positions with respect to the color image.

The database production can be divided into two stages:

I. Manufacturing 3D Masks: Spoofing attacks using 3D facial masks cannot become common because of high-cost associated with it. Recently 3D printing services seems growing but unfortunately making different mask attacks possible to face recognition systems. In Morpho database masks are manufactured with the help of 3D models of the valid user using 3D printer. It is practically impossible to take hold of 3D face model of a valid user due to the constraints existing in the remarkably advancing 3D scanner technologies. We can use the web service called ThatsMyFace.com to produce the masks for our database. The images are uploaded on the web and spoofs are created using hard resin mask or paper craft mask. In Morpho database the facial surface is complete while eyes and the nostrils are cut in 3DMAD.

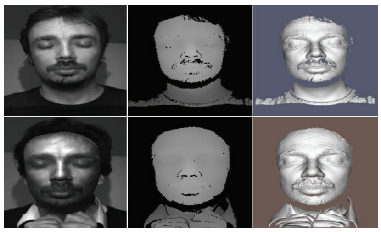


Fig.1 (a) real access from a user in grayscale (2D) texture

(b) Paper craft mask from 3DMAD is shown in top row; attacker Wearing a mask is shown in bottom row (2.5D and 3D model format).

(c) Example of hard resin masks from 3D MAD

II. Recording of databases: To record all the sample in database Microsoft kinect is used for both real access and mask attacks. The device is preferred over other conventional 2D-cameras since it is additionally provided with depth information that makes possible to analyze the vulnerability of 3D face recognition systems to mask attacks. Three sessions are carried out to collect the data samples from all users. In first two sessions only real access samples are collected and in third session mask

attacks are captured from a single attacker.

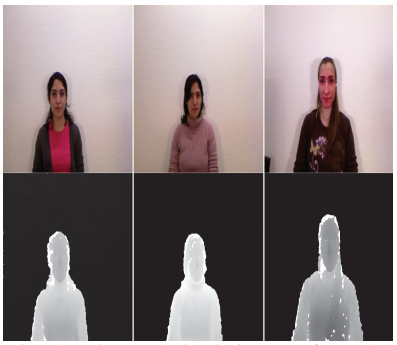


Fig. 2 color and depth images from 3DMAD. The first two in the upper row are real and third is mask.

III.ANALYSIS OF OUR APPROACH

The earlier proposed approach consisting of three steps will be analyzed to draw meaningful conclusions:

I. In the very first step the aim is to find a reasonable virtual training sample. A virtual training sample is reasonable if it has a proper difference with the corresponding original training sample. The mean of two of the original training samples is taken as a virtual training sample. Let a and b be two original training samples of same class (or face), then

$$C = (a + b) / 2$$

C is the corresponding synthesized virtual training sample. The difference between original training sample and the synthesized virtual training sample C is

$$Y = \|a-c\|_2 = \|(a-b) / 2\|_2$$

The value of Y is usually not too small or large because of varying facial expression, illumination and pose.

II. In this section analysis on BTSS is carried out which demonstrates that from the viewpoint of numerical computation it is not necessary in BTSS that all training samples should represent and classify the test sample. For high dimensional image data there is a theorem which states that "If a virtual training sample is a linear combination of the original training samples then we add the virtual training sample in a matrix A to form a new column, the new matrix has the same rank as the previous matrix". It gives the maximum possible number of the useful training samples.

III. The steps are closely linked with each other. In the first step all training samples are used to synthesize virtual training samples. Since virtual training samples reflect variations of the face so we select k useful training samples from BTSS and remaining training samples are discarded. In other words, in first step training samples that are far from the test sample are neglected and k training samples are exploited to represent test sample in the third step. These second and third steps are regarded as supervised sparse method.

Let us discuss anti-spoofing algorithms to detect 3D mask attacks, which become difficult with motion analysis or liveness detection methods. So we use texture analysis to differentiate between real faces and masks. Since human skin possess different optical characteristics from the masks.

Local Binary Patterns (LBP) is efficient texture operator used in counter measures against spoofing attacks. Proposed multi-scale LBP based feature vector is used against photo print attacks to detect 3D masks. For this purpose, it is applied separately on 2D and 2.5D images and classification rates achieved are 89.4% and 82.4% respectively. The two modes are later fused to achieve rates of 93.0% at feature level and 93.5% at score level. Let us compare various types of LBP operators for real face and mask classification using both 3DMAD and Morpho databases.

a. LBP based feature extraction: For each pixel the original LBP value is calculated comparing the value of pixel with its adjacent pixels in 3 × 3 neighborhood and from results 8-bit binary number is formed ($LBP_{8,i}$). Patterns with more than two bitwise transitions are eliminated, which is a common extension to the original operator. So the number of labels is reduced to 59 uniform patterns ($LBP_{8,1}^{u2}$).In whole image occurrences of the LBP labels are collected into histograms and are then considered as feature vectors for classification. The three types of LBP histograms $LBP_{8,1}^{u2}$,$LBP_{8,2}^{u2}$ and $LBP_{16,2}^{u2}$ are computed and concatenated which results in an enhanced feature histogram having length 833. Further three more extensions of LBP are included namely modified (mLBP), transitional (tLBP) and direction coded (dLBP). In mLBP average of the neighboring pixels are used for comparison instead of pixel value. In tLBP, comparison is drawn between two

consecutive neighboring pixels circularly in clock-wise direction. For dLBP the final result is 8-bit value which is formed by encoding of intensity variations only in four base directions of two bits.

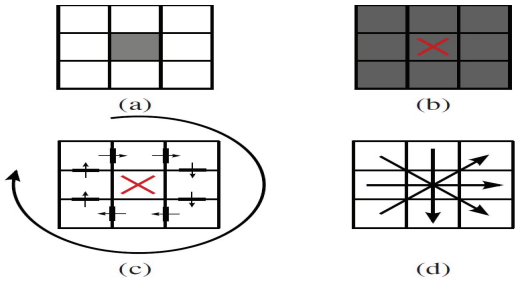


Fig.3 (a) conventional LBP, (b) mLBP (c) tLBP (d) dLBP

b. Feature classification: Since obtained histogram is a collection of extracted LBP codes, the classification can be carried by computing similarities of histogram. So calculation of two reference histograms is performed in the training set as the average of real access and mask attack samples. The comparison with the features extracted from test samples using x^2 metric, D_{real} and D_{mask} are two resulting distances obtained. D_{mask} - D_{real} is calculated to obtain the final result.

IV.EXPERIMENTS AND RESULTS

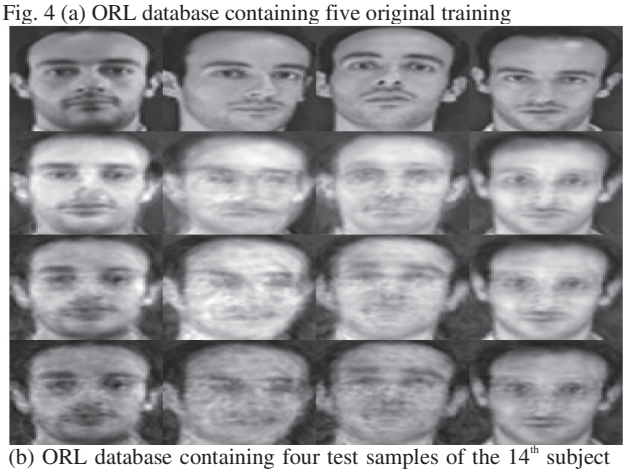
The experiments were conducted on different face databases like FERT, LFW, ORL, AR and GEORGIA TECH. The proposed approach for dealing uncertainty is compared with other methods based on appearance including K-nearest neighbor (KNN), linear discriminate analysis (LDA), principal component analysis (PCA), collaborative representation based classification (CRC) and some other. The regularization parameter θ in our approach is set to 0.01 for all experiments. Unconstrained face recognition problem is studied for designing LFW database, which is a collection of more than 13000 face images from web. Name of the person is labeled with each face pictured. Two or more distinct photos of 1680 people pictured were found in the dataset. Detection was only possible by Viola-Jones face detector which is the only constraint for these faces. 1251 images have been chosen from 86 people in our experiments. 11-20 images from each person which is manually cropped and resized to 32 × 32 pixels. Let us evaluate cases, in first case training samples were taken first seven images of each subject and test images were taken as remaining ones. In second case first eight images are used as training samples and the remaining ones as test sample.

Table I: Showing recognition accuracy and running of different methods in LWF database.

Experiment protocol	Case 1		Case 2	
	Recognition accuracy	Time (second)	Recognition accuracy	Time (second)
CRC	11.25%	0.0054	11.01%	0.0058
PCA	16.18%	0.0055	16.70%	0.0059
LDA	13.86%	0.0082	11.01%	0.0139
KNN(K=4)	16.18%	0.0010	17.23%	0.0011
SRC(l1_ls)	16.02%	21.668	16.34%	31.406
SRC(DALM)	18.49%	0.9611	19.36%	1.0616
SRC(DALM_fast)	12.84%	0.0183	13.49%	0.0192
SRC(FISTA)	13.69%	0.1594	16.01%	0.1687
SRC(HOMOTOPY)	14.02%	0.2002	15.09%	0.2021
SVM	31.12%	0.0045	31.43%	0.0058
LRC	31.59%	0.0753	33.21%	0.0871
Our approach	34.36%	0.0112	36.41%	0.0168

Let us compare our proposed approach with CRC and SRC (l1-ls). Five original training samples and corresponding ten synthesized virtual training samples from ORL database is shown in figure 4(a).

Fig. 4 (a) ORL database containing five original training



(b) ORL database containing four test samples of the 14th subject and different approaches are used to obtain corresponding representation results.

Difference in lighting and expression can be observed in virtual training samples and original training samples. From ORL database 4 test samples of the 14th subject and the corresponding results obtained through different approaches are shown in fig. 4(b). We can obtain the representation coefficients using eq. 4 and 5 discussed earlier. In fig. 4(b) results we can clearly see that the representation results obtained through our proposed approach in second row are comparatively better than third and fourth rows which are obtained by CRC and SRC respectively. Eq. 6 can be used to obtain the residual of each class

in our proposed approach. For CRC and SRC (l1-ls) residual of each class can be obtained by:

$$\|y - x_i p_i\|$$

Where test sample is represented by y. The matrix of the i_{th} class original training samples and the representation coefficients corresponding to them are respectively represented by x_i and p_i. We can observe the residual values are smallest for our approach which implies that it is easier to classify the test sample using our proposed approach. The representation coefficients have sparsity since in first column most of the residuals are equal to 1. This is induced by the scheme of representing and classifying the test sample by selecting and exploiting the useful training samples of BTSS. The recognition accuracy varies with the number of useful training samples k. When k becomes small the proposed approach tends to yield better performances. Generally, we set the range of k as (0.05)N – (0.2)N, where N is the number of original training samples. Moreover the occlusion of sunglasses is one of the most challenging task in face recognition accuracy. For this case, recognition accuracy of 90.83%, 0.41% is achieved in comparison to SRC (l1-ls).

For spoofing attacks, experiments are conducted on two types of databases:

I. Base line face recognition algorithms including Inter Session Variability (ISV) and Iterative Closest Point (ICP), a well established technique used for rigid registration of 3D surfaces, are used for the experimentation to access the success rates of spoofing attacks with 3D masks.

II. Experiments on anti- spoofing in which classification of mask attack or real face accuracy is measured through aforementioned counter measures.

Algorithms used in face recognition and anti-spoofing are summarized in the experimental in the diagram fig. (5). The results are displayed based on the image type and the decision taken. Face verification system takes a binary decision whether to accept or reject the input face. Micro-texture analysis algorithm is used for classification.

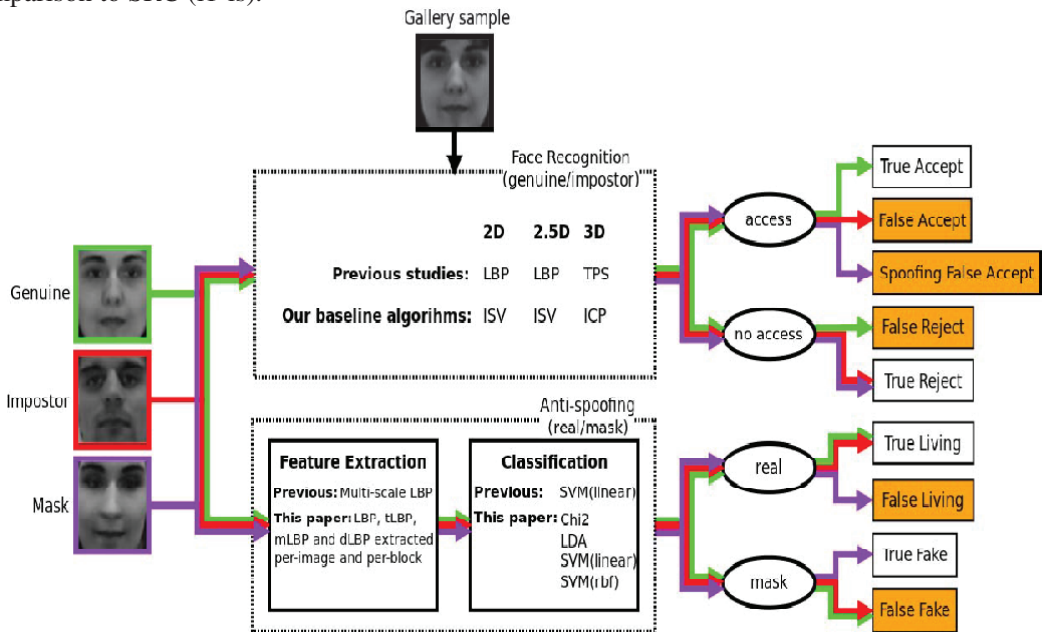


Fig.(5) A diagram contains face recognition and anti-spoofing modules with previously proposed algorithms for 3D mask attacks and the new ones studied in this paper.

V.CONCLUSION

The aim of this paper is to improve face recognition firstly by synthesizing virtual training sample from original training samples to go through various variations of face. Based on this, useful training samples are selected and exploited from BTSS for the representation and classification of test sample. This scheme improves the recognition accuracy by eliminating improper training samples.

On the other hand for biometric recognition systems spoofing attacks are still considered as big security issue since face is easily available and vulnerable to attacks. Majority of previous research in face spoofing is oriented towards preventing 2D attacks by displaying photos or video on mobile devices. However with the advancement in 3D technology face spoofing attacks are easier or somehow cheaper to carry by using 3D masks. In this paper we aim to continue our research in the domain of 3D mask attacks.

REFERENCES

[1] C. C. Aggarwal and P. S. Yu, ‘‘A survey of uncertain data algorithms and applications,’’ *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 5, pp. 609–623, May 2009.

[2] B. Qin, Y. N. Xia, S. Prabhakar, and Y. C. Tu, ‘‘A rule-based classification algorithm for uncertain data,’’ in *Proc. IEEE 25th Int. Conf. Data Eng.*, Apr. 2010, pp. 1415–1418.

[3] G. Cormode and A. McGregor, ‘‘Approximation algorithms for clustering uncertain data,’’ in *Proc. 27th PODS*, 2008, pp. 191–200.

[4] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, ‘‘Security evaluation of biometric authentication systems under real spoofing attacks,’’ *IET Biometrics*, vol. 1, no. 1, pp. 11–24, Mar. 2012

[5] K. Nixon, V. Aimale, and R. Rowe, ‘‘Spoof detection schemes,’’ in *Handbook of Biometrics*, A. Jain, P. Flynn, and A. Ross, Eds. New York, NY, USA: Springer-Verlag, 2008, pp. 403–423.

[6] J. Bai, T.-T. Ng, X. Gao, and Y.-Q. Shi, ‘‘Is physics-based liveness detection truly possible with a single image?’’ in *Proc. IEEE ISCAS*, May/Jun. 2010, pp. 3425–3428.

[7] J. Li, Y. Wang, T. Tan, and A. K. Jain, ‘‘Live face detection based on the analysis of fourier spectra,’’ *Proc. SPIE*, vol. 5404, pp. 296–303, Aug. 2004.

[8] J. Maatta, A. Hadid, and M. Pietikainen, ‘‘Face spoofing detection from single images using micro-texture analysis,’’ in Proc. IEEE IJCB, Oct. 2011, pp. 1–7.

[9] L. E. Ghaoui, G. R. G. Lanckriet, and G. Natsoulis, ‘‘Robust Classification with Interval Data,’’ Technical Report UCB/CSD-03-1279, Comput. Sci. Div., Univ. California, Berkeley, Oct. 2003.

[10] G. Lanckriet, L. Ghaoui, C. Bhattacharyya, and M. Jordan, ‘‘A robust minimax approach to classification,’’ J. Mach. Learn. Res., vol. 3, pp. 555–582, Dec. 2002.

[11] C. Bhattacharyya, K. Pannagadatta, and A. Smola, ‘‘A second order cone programming formulation for classifying missing data,’’ in Proc. NIPS, 2005, pp. 153–160.

[12] K. Kollreider, H. Fronthaler, and J. Bigun, ‘‘Evaluating liveness by face images and the structure tensor,’’ in Proc. IEEE Workshop Autom. Identificat. Adv. Technol., Oct. 2005, pp. 75–80.

[13] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay, ‘‘Moving face spoofing detection via 3D projective invariants,’’ in Proc. 5th IAPR ICB, Mar./Apr. 2012, pp. 73–78.

[14] G. Pan, L. Sun, Z.Wu, and S. Lao, ‘‘Eyeblink-based anti-spoofing in face recognition from a generic webcamera,’’ in Proc. IEEE ICCV, Oct. 2007,
pp. 1–8.

[15] N. Erdogmus and S. Marcel, ‘‘Spoofing 2D face recognition systems with 3D masks,’’ in Proc. Int. Conf. Biometrics Special Interest Group, 2013.