A Face Antispoofing Database with Diverse Attacks

Zhiwei Zhang[1], Junjie Yan[1], Sifei Liu[1], Zhen Lei[1,2], Dong Yi[1,2], Stan Z. Li[1,2*]
[1]CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences
[2]China Research and Development Center for Internet of Thing
{zwzhang,jjyan,sfliu,zlei,dyi,szli}@cbsr.ia.ac.cn

Abstract

Face antispoofing has now attracted intensive attention, aiming to assure the reliability of face biometrics. We notice that currently most of face anti-spoofing databases focus on data with little variations, which may limit the generalization performance of trained models since potential attacks in real world are probably more complex. In this paper we release a face anti-spoofing database which covers a diverse range of potential attack variations. Specifically, the database contains 50 genuine subjects, and fake faces are made from the high quality records of the genuine faces. Three imaging qualities are considered, namely the low quality, normal quality and high quality. Three fake face attacks are implemented, which include warped photo attack, cut photo attack and video attack. Therefore each subject contains 12 videos (3 genuine and 9 fake), and the final database contains 600 video clips. Test protocol is provided, which consists of 7 scenarios for a thorough evaluation from all possible aspects. A baseline algorithm is also given for comparison, which explores the high frequency information in the facial region to determine the liveness. We hope such a database can serve as an evaluation platform for future researches in the literature.

1. Introduction

Although face recognition (FR) has achieved great success during the past decades, little effort has been made to assure its security and reliability in the real world applications. It is now increasingly aware that existing FR systems are susceptible to fake face attacks, through which unauthorized attackers try to access illegal authorities by exhibiting fake faces of an authorized client. Serious consequences may occur if these attacks succeed, yet unfortunately there still lack effective anti-spoofing techniques.

Nowadays attackers can obtain a client's face images by using portable digital cameras or simply downloading from the Internet, and fake faces can be easily produced, for example, printing photos or showing videos on a laptop. Fake faces like photos and video playbacks are not only easy to implement but also usually quite effective in spoofing a FR system [1], and has become the main concern [1] in the literature as shown in Section. 2. Actually in real world applications, FR systems may encounter various fake face attacks, and an excellent anti-spoofing algorithm should perform robust in unpredicted situations.

However, as we review the recent development in Section. 2, we find that current researches only concentrate on fake face with little variations. For example, in NUAA database [2], only photo attacks are considered which are warped as attack; in Idiap database [3] photos are either fixed or hand-hold. A more complex case appears in [4] where the eye regions are cut off the photo so that attackers can hide behind and blink; however, the authors did not release the database and the data amount is relatively small (only 13 subjects involved). Another shortcoming in variation is the quality of attacks: in [5, 2, 6], algorithms extract the high frequency information (e.g. texture) to detect liveness. But as these high frequency information highly depends on the image quality, how will they perform on good quality and bad quality images? Do their algorithms generalize well? We can see that due to the lack of variational data, many questions remain unanswered. Therefore we cannot predict their performance in real world applications, because practical attacks are probably not limited to one single type as in previous researches.

The above observation motivates us to release a comprehensive database to serve as an evaluation platform for the face anti-spoofing issue. The database consists of 50 subjects, and we carefully design 3 kinds of imaging qualities and three kinds of attacks, which yields 12 videos for a single subject. A test protocol including 7 scenarios is provided, which analyze the effect of different imaging qualities and attack types thoroughly. Compared with previous

*Stan Z. Li is the corresponding author

[1]Mask is also a good choice, but usually it's too expensive to produce client-like masks. So the massive usage of masks rarely appears in the literature.

978-1-4673-0397-2/12/$31.00 ©2012 IEEE

databases, our database has a much wider coverage of data variation as shown in Section. 3. We further design a baseline algorithm to give a preliminary study. For fake faces, the reproduction process such as printing or displaying will inevitably degrades the quality of facial texture. So the high frequency information may be a strong proof of liveness. We use multiple DoG filters to extract the high frequency information and exclude the low frequency information and noise. Then a SVM classifier is trained on the filtered image, which outputs the final decision.

The paper is organized as follows: in Section 2 we review the mainstream anti-spoofing algorithms and fake face data used in their works; in Section 3 we explain our database in details; in Section 4 we give the test protocol and the baseline approach, and in Section 5 we conclude this paper.

2. Literature Survey

Basically there are two main categories of face anti-spoofing techniques, the facial motion detection category and facial texture analysis category. Facial motion detection techniques expect subjects to exhibit specific facial motion, the detection of which determines the liveness. Facial texture analysis techniques believe that fake faces probably lack some high frequency information during the reproduction process, and by analyzing and learning the facial texture information, genuine and fake faces can be classified properly. Here the term "texture" represents the high frequency details in face images, and without ambiguity, we treat "texture" equally with "high frequency information".

It is important to mention that multi-spectral [7, 8, 9] and multi-modality [10, 11] based anti-spoofing methods also perform quite well by exploring other liveness clues such as face images under nonvisible illuminations or voice information. However, as they all require extra hardware devices which may be absent in most of existing FR systems, their applications are limited. So in this paper we restrict our concern only on the unimodal face biometrics under visible illumination just as the NUAA [2] and Idiap [3] database do.

2.1. Facial Motion Detection Category

Several kinds of facial motions can be utilized in this category, such as eye blinking, mouth movement, head rotation, etc. Eye blinking is perhaps the most widely used facial motion due to its naturalness and simplicity. In [4], optical flow is used to measure the blink motion. In [12, 13], authors explore the appearance features to learn an eye state classifier. The state variation represents the blink motion. Head rotation [14] and mouth movement [10] are similarly measured. [4] combines eye blinking and head rotation together.

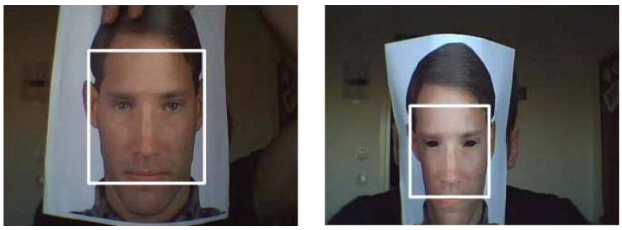

Figure 1. Fake faces used in [4], left is warped photo attack, and the right is cut photo attack. The figure is reproduced from [4].

Despite their intent to anti-spoof fake faces, unfortunately no fake face data are released. Among the above works, [12] used warped photos as attack; [4] designed several fake faces, which include warped photo attack and cut photo attack. Cut photo attack means that certain parts of the photo are cut off so that attackers can hide behind and exhibit facial motion through the hole. Furthermore, the amount of fake faces used in their experiments are relatively small. Some fake face samples can be seen in Fig. 1.

2.2. Facial Texture Analysis Category

Methods of this category believe that during the reproduction process from the genuine face to fake ones, high frequency information will be lost. In [5] Fourier transform is utilized to extract the high frequency information, and the target face image is judged fake if its energy percentage of high frequency is lower than a certain threshold. In [2] the authors use DoG and LTV algorithms to extract high frequency information from the captured images, and the final model is learned by a complex bilinear sparse low rank logistic regression model. In [6] a more simple but also more powerful LBP+SVM method (named Micro-Texture Analysis, MTA) is proposed, and they achieve very amazing results both on the NUAA [2] and Idiap [3] database.

Figure 2. The first row is the warped photo attack in [2], and the second row is the fake faces in [3], photos either fixed or hand hold. Notice that very obvious print defects can be seen.

Similarly [20] also utilize micro-textures and SVM to detect liveness.

In NUAA database, 12641 still images are captured from 15 subjects. Warped photos of three sizes ($6.8cm \times 10.2, 8.9cm \times 12.7$ and A4 size) are taken as attacks. Idiap database, which is used for the IJCB'11 anti-spoofing contest, contains 50 subjects. Each subject is captured twice in different environment, and corresponding photo attacks are implemented by either fixing or hand holding the photos. For each of the above cases, two sessions are recorded, yielding a total 400 videos. Some examples are shown in Fig. 2.

3. The Collected database

In this section, we will describe the collected face anti-spoofing database. We begin with the introduction of the imaging devices as well as the imaging quality, then we introduce how our genuine faces and fake faces data are collected. A statistical comparison with other databases is given at last, which shows our evident advantage in data collection.

3.1. Imaging Quality

Imaging quality is an issue which hasn't been discussed before in face anti-spoofing. Generally speaking, the performance of an algorithm depends on the image quality to some extent. Furthermore, the requirement on imaging quality also determines the imaging devices to collect data. For the sake of simplicity, here we empirically define quality as the preservation of facial textures, by which we pay more attention to the perceptual feeling rather than strict quantitative measures.

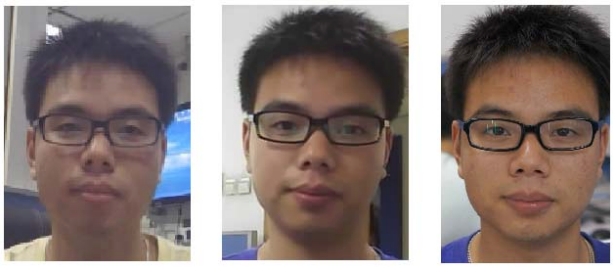

Figure 3. Videos of low, normal and high quality. Only face regions are shown. Zoom in the image, and more obvious difference can be seen.

We use three different cameras to record the data of different qualities. The low quality video is captured by a long-time-used USB camera since long time usage will always degrade the imaging quality. The image height and width for low quality videos is 640 and 480. The normal quality video is captured by a newly bought USB camera which can keep the original imaging quality well. The image height and width is 480 and 640. For the high quality videos, we use a high resolution Sony NEX-5 camera for recording, whose maximum resolution is up to 1920×1080. We take one image with the maximum 1920×1080 resolution for each subject as the material to make photos. For genuine face videos and video attacks, however, such a high resolution video is too heavy a burden for both saving and calculation. Therefore we only crop a 1280×720 image patch which contain faces to form the final videos. By doing so, we reduce the saving and calculation burden while maximizing the quality we can achieve. Some examples can be seen in Fig. 3, in which only the face parts are shown. A complete video set can be seen in Fig. 6.

3.2. Genuine faces

50 subjects are collected to form the genuine faces. All subjects are captured in natural scenes with no artificial environment unification. During recording, subjects are required to exhibit blinking behavior rather than keeping still. We argue that facial motion is a crucial liveness clue for anti-spoofing, and it is necessary to provide them just like in a challenge-response strategy used in facial motion detection methods. The motion type of blink is chosen because it is more natural and user-friendly than other motion types such as head movement and mouth movement. A blink process can be seen in Fig. 4.

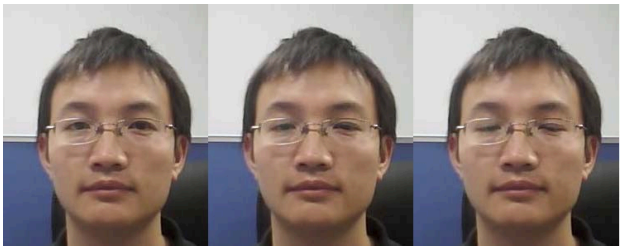

Figure 4. A blink process.

3.3. Fake faces

Fake faces are the key ingredients in our database. Specifically, we design the following three kinds of fake face attacks.

Warped photo attack. As mentioned above, we use a Sony NEX-5 camera to record a 1920×1080 image and a 1280×720 video for every subject. We use this high resolution image to print the photos, and these photos are printed on the copper paper, which has much higher quality than normal A4 printing paper. In a warped photo attack, the attacker deliberately warps an intact photo, trying to simulate facial motion. Intact means there is no cut-off region in the photo, in contrast with the cut photo attack below. Our warped photo attacks are much similar with those in Fig. 2.

Cut photo attack. The photos mentioned above are then used for the cut photo attacks. As we require subjects to ex-

Figure 5. Different fake face attacks: left, attacker hides behind the cut photo and blink;(2)middle, another intact photo is up-down moved behind the cut one;(3)right, is the video attack.

hibit blink behavior, here the eye regions are cut off. An attacker hide behind and exhibit blinking through the holes as shown in Fig. 5. Another possible implementation is proposed in [4] that one intact photo is tightly placed behind the cut photo, and by moving the photo behind, blinking can be simulated. In this database both the two implementations exist.

Video attack. In this case, the high resolution genuine videos are displayed using an iPad. Notice that limited by the iPad screen resolution, the original high resolution videos (1280×720) will be inevitably downsized by the device. One example can be seen in Fig. 5.

3.4. Comparison with Previous Databases

In Table. 1 we give a statistical comparison with previous NUAA and Idiap databases, which are the only available databases to face anti-spoofing researchers. NUAA database uses photos of different sizes as attacks as shown in Table. 1. In contrast, in Idiap and our database, only A4-sized photos are used to maximize the preservation of facial textures. Furthermore, we print photos on copper paper, which has even higher quality than common A4 printing paper. In Idiap database, intact photos are both fixed and hand hold as attacks. By comparison we can see that our database has an obvious advantage not only in data variation but also in the data amount. Fig. 6 shows a complete video set for a single individual. The whole database can be downloaded from http://www.cbsr.ia.ac.cn/english/FaceAntiSpoof%20Databases.asp

Table 1. Comparison with previous databases

	NUAA[2]	Idiap[3]	Ours
#Subject	15	50	50
Data type	Still image	Video	Video
#Data	12641	400	600
Attack	Warped photo of size 1. $6.8 \times 10.2cm$ 2. $8.9 \times 12.7cm$ 3. A4 size	Intact photo 1. Fixed 2. Hand Hold	1.Warped photo 2.Cut photo 3.Video playback
#Quality	1	1	3

4. Protocol and Baseline

In this section we give the test protocol first. Then we introduce a simple algorithm to serve as the baseline. Finally we analyze the experimental results.

4.1. Test protocol

Our database mainly considers the possible effect of imaging quality and various fake faces, therefore it is necessary to provide a thorough analysis by verifying algorithms under different scenarios. Specifically, we design a test protocol which consists of 7 scenarios, each of which involves only certain samples. Notations L,N,H represents the low quality, normal quality and high quality, see Fig. 6.

- **Quality Test.** The three imaging qualities are considered explicitly. Specifically, the samples used are:
  - **1.** Low quality test: only use {L1,L2,L3,L4}.
  - **2.** Normal quality test: only use {N1,N2,N3,N4}.
  - **3.** High quality test: only use {H1,H2,H3,H4}.
- **Fake Face Test.** Similarly, the three fake face types are considered explicitly. Samples used are:
  - **4.** Warped photo attack test: {L1,N1,H1,L2,N2,H2}.
  - **5.** Cut photo attack test: {L1,N1,H1,L3,N3,H3}.
  - **6.** Video attack test: {L1,N1,H1,L4,N4,H4}.
- **Overall Test.** In this case, all data are combined together to give a general and overall evaluation.
  - **7.** Overall test: all the videos are used.

The whole database has been split into the training set (containing 20 subjects) and the testing set (containing 30 subjects) already. For each of the above 7 scenarios, the corresponding data are to be selected from the training and testing set for model training and accuracy testing.

Similar with [3], Detection-Error Trade-off (DET) curves [19] are utilized to evaluate the anti-spoofing accuracy. From DET curves, the point where FAR equals FRR is located, and the corresponding value, which is called the Equal Error Rate (EER), should also be reported. For any evaluating algorithm, 7 DET curves and 7 EER results should be reported corresponding to the above 7 scenarios. Examples can be seen in the following section for the baseline algorithm. The DET plotting code can be downloaded from http://www.itl.nist.gov/iad/mig/tools/

4.2. Baseline algorithm

Fake face images can be viewed as the genuine face images post-processed by the reproduction process. It is known that the printing, imaging and displaying process can introduce distortions such as blur and aliasing [2, 18]. Then the captured fake face images should possess lower quality

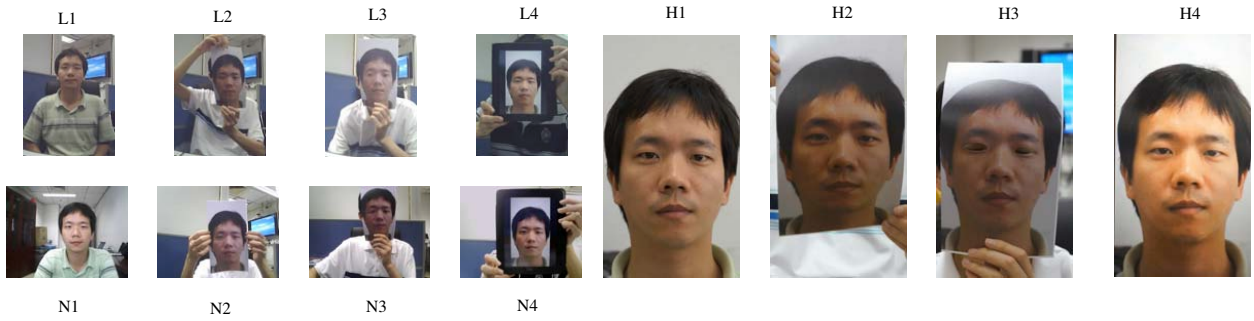

Figure 6. One complete video set for a subject. The left top four images represent the low quality videos, the left bottom are the normal quality videos, and the right are the high quality videos. For each quality, from left to right are genuine, warped photo attack, cut photo attack and video attack.

than the genuine faces. Based on this intuitive idea, here we construct a simple baseline algorithm.

We use multiple Difference of Gaussian (DoG) filters to extract the high frequency information from the face images, which is treated as the liveness clue. As we have no prior knowledge about which frequency is most discriminative, we adopt multiple DoG filters to form a redundant feature set. The low frequency information and the noises can also be excluded by properly setting the Gaussian σ. Specifically, let σ_1 represents the inner Gaussian variance, and σ_2 the outer one. Then four DoG filters are considered here: $\sigma_1 = 0.5, \sigma_2 = 1$; $\sigma_1 = 1, \sigma_2 = 1.5$; $\sigma_1 = 1.5, \sigma_2 = 2$; and $\sigma_1 = 1, \sigma_2 = 2$. Then the concatenated filtered images are taken into SVM to train a final classifier.

As in this database each data is a video sequence, we randomly select 30 faces from the videos to train the classifier. Then for test we also randomly select 30 faces and identify them either as "genuine" or as "fake". The final label for the video are determined by averaging the 30 image scores. Moreover, the filtered images are down-sampled to reduce the feature dimension. The pipeline of the algorithm is shown in Fig. 7.

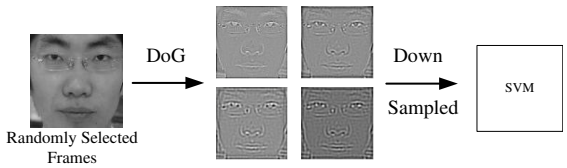

Figure 7. Baseline algorithm.

4.3. Experiment Results & Analysis

In Fig. 8 we show seven DET curves for the above designed scenarios. Then in Table.2 the EER are given from the DET curves where FAR equals FRR.

Firstly we examine the effect of image quality toward anti-spoofing accuracy. We can see from Table.2 that when captured in very high quality, the EER is worst. We believe

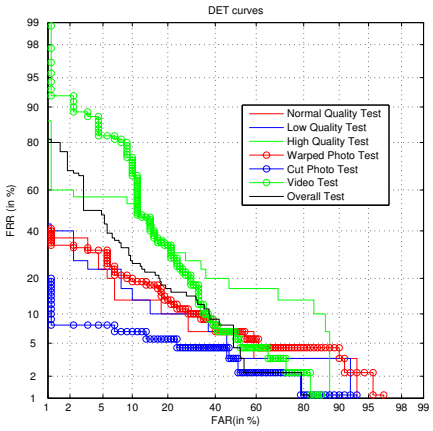

Figure 8. DET curves.

Table 2. Equal Error Rate

scenario	1	2	3	4	5	6	7
EER	0.13	0.13	0.26	0.16	0.06	0.24	0.17

that obvious textures can also be extracted from fake faces captured in high quality. In that case, the assumption that the quality of fake faces tend to be lower than that of the genuine faces will be weakened, making the classification more difficult.

Then we investigate the different fake face types. From Table.2 we can see that the performance on cut photo attack is the best while the video attack is the worst. Cut photo attack is easy to conquer, probably because there exist sharp edges in similar eye regions, making the negative samples less variational. It is also reasonable that it is very difficult to anti-spoof video attacks, because videos can naturally exhibit a live face, especially when the quality can be guaranteed. More powerful anti-spoofing methods have to be designed in future.

The EER for overall test is 0.17, which approximately falls between the best (0.06) and the worst (0.26) like a trade-off among different attack scenarios. This accuracy

is still unsatisfactory for real world applications, and there is much improvement to be achieved.

From the above experiments we can see that both imaging quality and fake face types can significantly influent the anti-spoofing accuracy. The quality test shows that different imaging quality can affect the inter-class difference. While low and normal qualities achieve the same EER, high quality is much worse, implying that it may be not always good to pursue high quality in imaging. The fake face test shows that a successful attack should preserve the face appearance as much as possible. The existence of obvious cutting edge in cut photo attacks can help anti-spoofing, while in warped and video attacks there exist no such similar ingenuity evidence. And finally the overall test implies that building a universal anti-spoofing model which covers many factors is relatively harder than concentrating on some fewer factors. That' why we release this database, because we believe in real world situations, all these cases may occur. By providing these data of large variation, we do hope our database can assist future research in the issue.

5. Conclusion

In this paper we release a face anti-spoofing database with diverse attacks to serve as an evaluation platform in the literature. The database contains 50 genuine subjects, and the fake faces are produced from the high quality records of the genuine faces. Three imaging qualities and three kinds of fake face attacks are included. We also design a test protocol which consists of 7 scenarios to provide a thorough analysis of different factors which may affect the anti-spoofing accuracy. We further design a DoG+SVM algorithm to explore high frequency information to classify genuine and fake faces, which serves as the baseline algorithm. To our knowledge, this is the most comprehensive database to date in the literature.

6. Acknowledgement

This work was supported by the Chinese National Natural Science Foundation Project #61070146, #61105023, #61103156, #61105037, National IoT R&D Project #2150510, and European Union FP7 Project #257289 (TABULA RASA http://www.tabularasa-euproject.org), and AuthenMetric R&D Funds.

References

[1] N.M. Duc and B.Q. Minh. Your face is not your password. Black Hat Conference, 2009.

[2] X. Tan, Y. Li, J. Liu and L. Jiang. Face Liveness Detection from A Single Image with Sparse Low Rank Bilinear Discriminative Model. ECCV, 2010.

[3] A. Anjos and S. Marcel. Counter-Measures to Photo Attacks in Face Recognition: a public database and a baseline. IJCB'11.

[4] K. Kollreider, H. Fronthaler and J. Bigun. Verifying Liveness by Multiple Experts in Face Biometrics. CVPR Workshop, 2008.

[5] J. Li, Y. Wang, T. Tan and A.K.Jain. Live Face Detection Based on the Analysis of Fourier Spectra. SPIE Biometric Technology for Human Identification, 1999.

[6] M. P. Jukka, Abdenour Hadid and Matti Pietik¨ainen. Face spoofing detection from single images using micro-texture analysis. IJCB, 2011.

[7] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In Computer Vision Beyond the Visible Spectrum: Methods and Applications, 2000.

[8] Zhiwei Zhang, Dong Yi, Zhen Lei and Stan Z. Li. Face liveness detection by learning multispectral reflectance distributions. In Automatic Face and Gesture Recognition, 2011.

[9] Y. Kim, J. Na, S. Yoon, and J. Yi. Masked fake face detection using radiance measurements, Journal of the Optical Society of America A, vol. 26, no. 4, 2009.

[10] G. Chetty and M. Wagner. Liveness Verification in Audio-Video Speaker Authentication. In 10th Australian Int. Conference on Speech Science and Technology, December, 2004.

[11] R. Frischholz, U. Dieckmann, BioID: A Multimodal Biometric Identification System, IEEE Computer,2000.

[12] G. Pan, L. Sun, Z. Wu and S. Lao, Eyeblink-based Anti-Spoofing in Face Recognition from a GenericWebcamera, ICCV, 2007.

[13] C. Xu, Y. Zheng and Z. Wang. Eye States Detection by Boosting Local Binary Pattern Histograms. ICIP, 2008.

[14] K. Kollreider, H. Fronthaler and J. Bigun, Evaluating Liveness by Face Images and the Structure Tensor, Fourth IEEE Workshop on Automatic Identification Advanced Technologies, October, 2005.

[15] P. Viola and M. J. Jones. Robust real-time face detection. In IJCV, 2004.

[16] IJCB'11 Competition on counter measures to 2D facial spoofing attacks. http://www.cse.nd.edu/IJCB_11/

[17] C. Chang and C. Lin , LIBSVM, http://www.csie.ntu.edu.tw/cjlin/libsvm/

[18] N. Joshi, R. Szeliski and D. J. Kriegman. PSF Estimation using Sharp Edge Prediction. CVPR, 2008.

[19] A. Martin, G. Doddington, T. Kamm, M. Ordowski and M. Przybocki. The DET Curve in Assessment of Detection Task Performance. European Conference on Speech Communication and Technology, 1997.

[20] J. Bai, T. Ng, X. Gao and Y. Shi. Is physics-based liveness detection truly possible with a single image? International Symposium on Circuits and Systems, 2010.