Face Recognition Performance Comparison of Fake Faces with Real Faces in Relation to Lighting

Mi-Young Cho* and Young-Sook Jeong
Electronics and Telecommunications Research Institute
218 Gajeongno, Yuseong-gu, Daejeon, 305-700, KOREA
{mycho, ysjeong}@etri.re.kr

Abstract

Face Recognition is widely used in security systems, such as surveillance, gate control systems, and guard robots, due to their user friendliness and convenience compared to other biometric approaches. Secure face recognition systems require advanced technology for face liveness detection, which can identify whether a face belongs to a real client or a portrait. However, with the development of display devices and technology, the tools and skills for carrying out spoofing attacks with images and videos have gradually evolved. In this paper, we compare real faces with high-definition facial videos from LED display devices, and present the changes in face recognition performance according to lighting direction.

Keywords: fake faces, face recognition, spoofing

1 Introduction

As interest in face recognition technology has grown, it has been applied in various secure applications, such as surveillance, gate control systems, and guard service robots. It is well known that face recognition systems are used to validate whether the subject matches a previously recorded facial image for identifying the users, while facial occlusion detection systems check the presence of any recognizable facial images. However, these techniques can be highly vulnerable to spoofing attacks, in which an impostor tries to bypass the face recognition system[2]. Generally, illegal attacks are performed with printed photographs, screen images or videos, ultra-realistic face masks, or 3D models of authorized clients. Among these types of attacks, printed photographs or screen images are easily used. Recently, attacks using images and videos have increased due to the development of high-resolution image capture and display devices.

Existing fake face detection approaches can be mainly categorized into three groups: texture-based, motion-based, and life sign-based. Texture-based approaches use unique texture patterns with 2D still images, such as print failures and overall image blur[13], [15]. However, very complex paper and printing textures can occur, and the systems for texture analysis must be sufficiently robust to process different texture patterns, which require the existence of a very diverse dataset. It is also possible for attacks to be performed using a photo displayed on a screen, which will produce very low-texture information. Motion-based approaches mainly use optical flow from image sequences to detect fake faces[1]. Motion analysis is helpful for avoiding dependence on certain texture patterns. However, motion analysis can present problems when there is low-motion information. Motion analysis requires a video feature and can also fail when spoof attacks are performed using 3D face models. Life sign-based approaches aim at detecting the physiological responses of faces. This response can be represented by biometric motions,

Journal of Internet Services and Information Security (JISIS), volume: 4, number: 4 (November 2014), pp. 82-90
*Corresponding author: Electronics and Telecommunications Research Institute, Tel: +82-42-860-6453, Web: http://www.etri.re.kr/etri/main/index.etri

such as eye blinking, head rotation, mouth movements, and multiple movements[14]. However, tools and skills for disguising motions have gradually evolved, such as masks and camouflages. To tackle these problems, researchers have considered using extra sensors as well as visual cameras. Andrea et al. [10] distinguished between 2D photos and real human faces using 3D vector scans. Sooyeon Kim et al. [8] compared two sequences of images using the in-focus and out-of-focus functions of a camera.

Despite the success of the above methods in some cases, non-intrusive methods without additional devices and human involvement are preferable in practice, since they can be easily integrated into an existing face recognition system, where usually only a generic webcam is equipped. Their methods are also disadvantageous in terms of cost and commercialization. In this paper, we focus on a generic face recognition system, where a webcam is equipped. We compare real faces with fake faces created by a high-resolution camera, and consider the problem of fake face detection under poor lighting conditions.

The paper is organized as follows: in Section 2, we introduce related works about fake face detection. Section 3 gives an overview of the experiment. Section 4 describes how to construct the imposter video database. In section 5, we show and analyze the experimental results. Section 6 concludes this paper.

2 Related Works

Approaches for liveness detection are categorized according to the following indicators: texture, motion, and life signs. The texture approach uses texture patterns, such as print failures and overall image blur. This approach assumes that fake faces are printed on paper. Gahyun Kim et al. [7] propose that 2D paper masks can be differentiated from real face images by their shape and level of detail. They used a power spectrum-based method for the frequency analysis due to two reasons: 1) the differences in low-frequency regions related to the lighting components generated by overall face shape, and 2) the differences in detail information between real faces and masks in the high-frequency region. The texture information is recorded, as the images taken from the 2D objects tend to suffer from a loss of texture information compared to the images taken from the 3D objects. This approach can be easily implemented without special equipment, and it does not require the user's cooperation. However, the system must be sophisticated enough to process different texture patterns, because diverse papers and printing textures can occur. Attacks can also be performed using a photo displayed on a screen.

The second approach takes advantage of motion indicators, which use differences in motion patterns between 3D objects and 2D planes. Kollreider et al. [9] combined face parts detection with an estimation of optical flow for face liveness detection. This approach uses a model-based Gabor decomposition and SVM(Support Vector Machine) for detection of face parts. This method is based on the basic assumption that a 3D face generates a 2D motion that is higher at central face regions than at the outer face regions, such as the ears. Therefore, facial parts which are farther away move differently to parts that are nearer to the camera. However, a photograph generates constant motion for different face regions. Motion analysis is helpful for avoiding dependence on certain texture patterns, and it does not require the user's cooperation. However, motion analysis can present problems when there is low-motion activity.

Finally, life signs are categorized into two types depending on interaction from the user. The first assumes a certain known interaction from the user. In this situation, the user needs to perform a certain task to verify the liveness of their face image as a challenge response or a motion password. The second category focuses on certain movements without user interaction, such as eye blinking, and considers those movements as a sign of life and therefore belonging to a real face. Jee et al. [6] introduced the technique based on an analysis of the movement of eyes with an embedded face recognition system. They proposed a method for detecting eyes in sequential input images, and calculated the variations in each eye region to determine whether the input face is real or not. Sun et al. [11] introduced the blinking-based approach for liveness detection using Conditional Random Fields (CRFs). They used CRFs to

model blinking activities, which were represented by image sequences of images with closed and non-closed states. Life signs are very hard to spoof with 2D images and 3D sculptures. This approach is also independent of textures, but may require user cooperation, and depends on face part detection.

In the past, it was difficult to differentiate real faces from similar facial images, due to the insufficient technology of display devices. However, it is now hard to distinguish real faces from fake faces created with high-definition display devices, such as UHD(Ultra High Definition) or HD(High Definition) monitors, with the unaided eye. In this paper, we compare real face images with fake facial videos created with high-definition display devices. In particular, we consider the performance gap between real faces and fake faces in relation to lighting conditions.

3 Overview

As mentioned in Section 2, it is easy to deceive face recognition systems with facial images from screens or monitors without any additional devices. Recently, with the improved color gamut of display devices, recognition algorithms are not able to differentiate "live" faces from "not live" faces, which is a major issue in the security field. Facial videos were used considering the natural movement of users under similar with real service environment. The experiment overview is as follows.

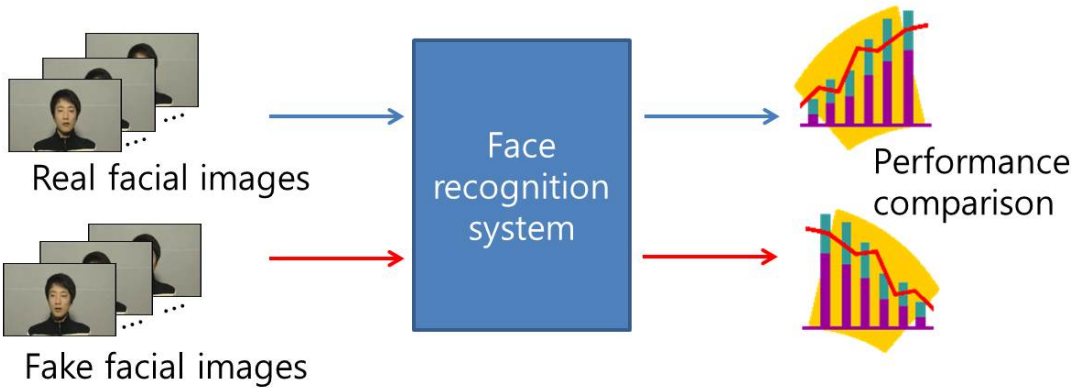


Figure 1: Experiment overview

A general web camera (Logitech HD Pro webcam C910) instead of a camera with a face recognition system was used to capture real facial movements and facial videos from a display device. A high-resolution LED or OLED monitor was used as a display device to provide output similar to that of a real face. The fake facial images were synchronized with the captured images of a real face, and both types of video data were then input into the face recognition system. Face recognition performances were then compared using detection rate (DR) and recognition rate (RR) as performance indicators.

4 Fake Facial DB

There are many fake face databases for research on liveness face detection. The NUAA [12] photograph database is collected using general cheap webcams; high-definition photos were taken for 15 subjects using a digital camera, and then developed into photos. To collect imposter photos, the photos were moved, rotated, or bent horizontally or vertically. Unlike the NUAA DB, the CASIA Face anti-spoofing database [5] mainly focuses on the variety of collected data to provide a comprehensive collection, including not only warped or cut photo attacks, but also video attacks. The database also contains three

imaging qualities; low quality, normal quality and high quality. Last, the Replay-Attack database [3] by Idiap Research Institute consists of 1300 video clips of photo and video attack attempts on 50 clients, taken under different lighting conditions.

Even though the existing fake face databases include video attacks, the definition was not sufficiently high to display through UHD monitors. Therefore, ultra-high resolution facial videos are needed to compare real faces with similar images. To capture facial videos, we covered the windows in a small studio room with a dark curtain to eliminate the influence of outside light. To obtain subject images under various lighting conditions, nine lights were used as directional light sources. The locations of lights are shown in Figure 2.

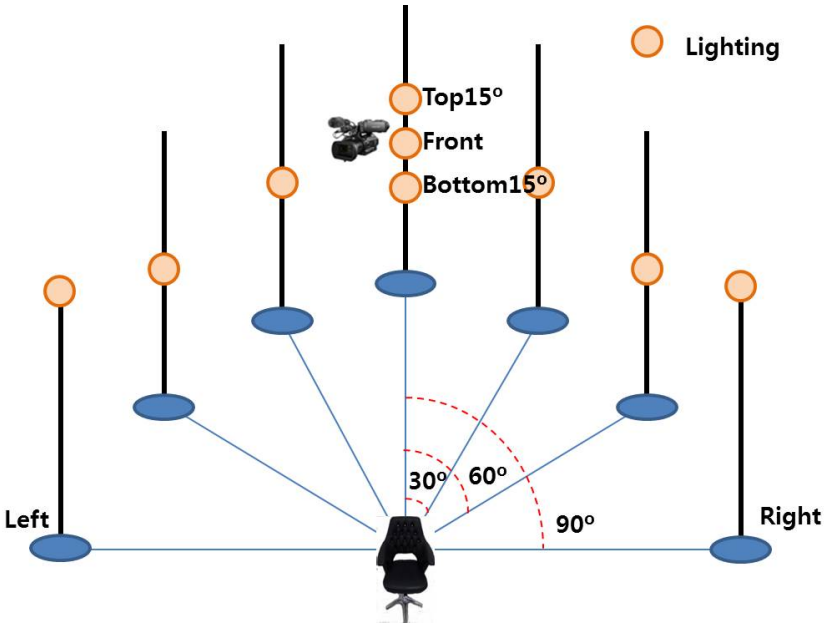


Figure 2: Capturing environment

We took ultra-high definition videos using a 2K broadcasting camera so that the face area took up at least two thirds of the whole area of the sequences. The height of the camera was fixed, and we controlled the height of the chair depending on the subject's height. We asked each subject to sit on a chair, look into the camera, and perform natural movements, such as speaking or smiling. We then captured the images with the camera, turning each light on and/or off. There were 60 subjects in total.

Compared with other databases, our database is more challenging, since it includes the captured images under various lighting conditions. A web camera and a broadcasting camera were used at the same time to capture the images of a real face to compare the performance.

To test their strength against video imposters, the facial videos captured with the broadcasting camera were displayed on a high-definition monitor, and then re-captured via a web camera. We displayed the video on a 50-inch LED monitor to provide an output similar to a real face. Before the capturing, the UHD monitor was calibrated using the monitor hood and calibration equipment. The International Commission on Illumination (CIE) recommends a gamma of 2.2, a brightness of 120cd–250cd, and a color temperature of 6500K to ensure proper monitor output [4]. The procedure for the fake face DB construction is shown in Figure 4. The web camera used was the same as the camera used for capturing real faces.

Each video clip was captured with 30fps and a size of 864 $\times$ 480 pixels for each configuration, lasting about 5 seconds-the same as the real facial videos. Figure 5 shows a histogram of real face and

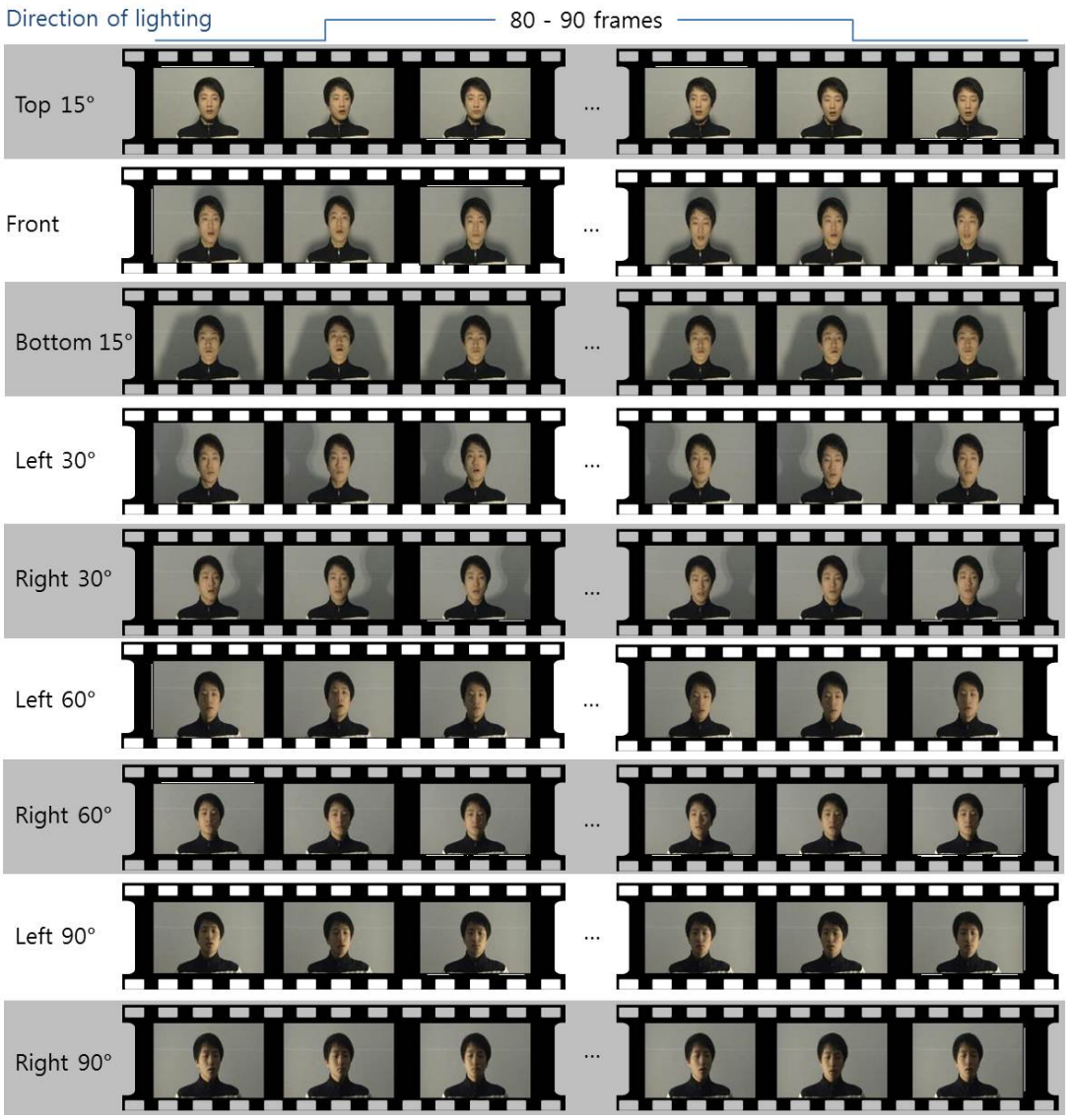


Figure 3: Capturing high-definition facial videos

LED fake face images.

5 Experiment Results

This experiment is not intended to test the performance of the face recognition system, but to compare face recognition performance between real facial and fake facial videos. In particular, we focused on changes in face recognition performance according to lighting direction and the features of fake facial videos. We tested 540 video clips, which were captured under nine different lighting directions for each subject. Before the performance comparison experiment, we synchronized the real- and the fake-facial videos.

We tested three face recognition engines used for commercial purposes. They have a different coverage range and performance for lighting direction. If the lighting coverage range of the engine is wide in general, it is regarded as a good recognizer. We used real facial images, and DR and RR were calculated

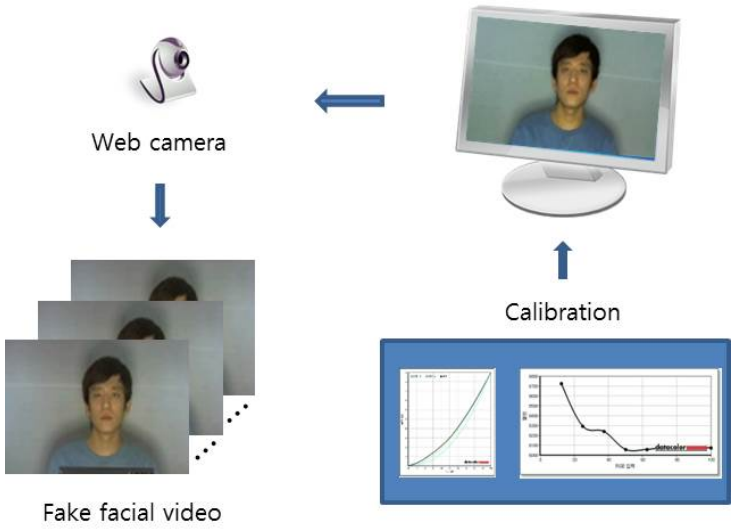


Figure 4: Fake face capturing procedure

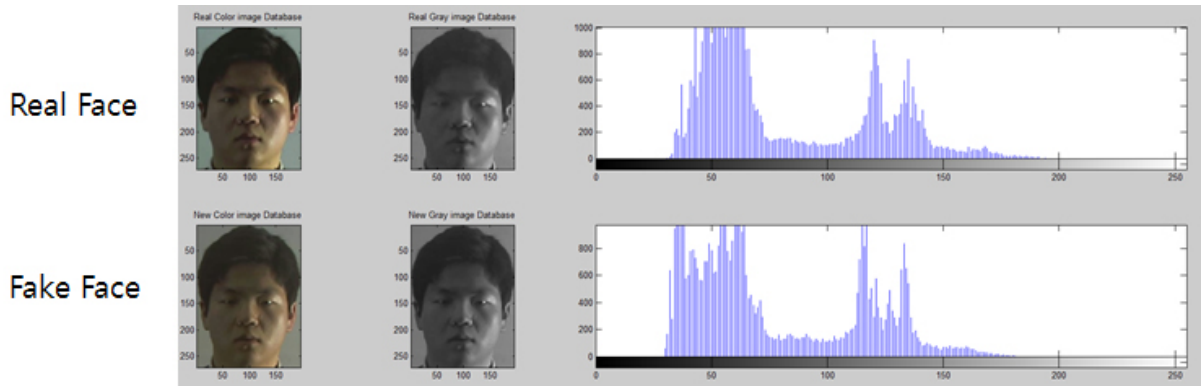


Figure 5: Histogram of sample real face and fake face images

to measure the performance of the real and fake LED facial videos.

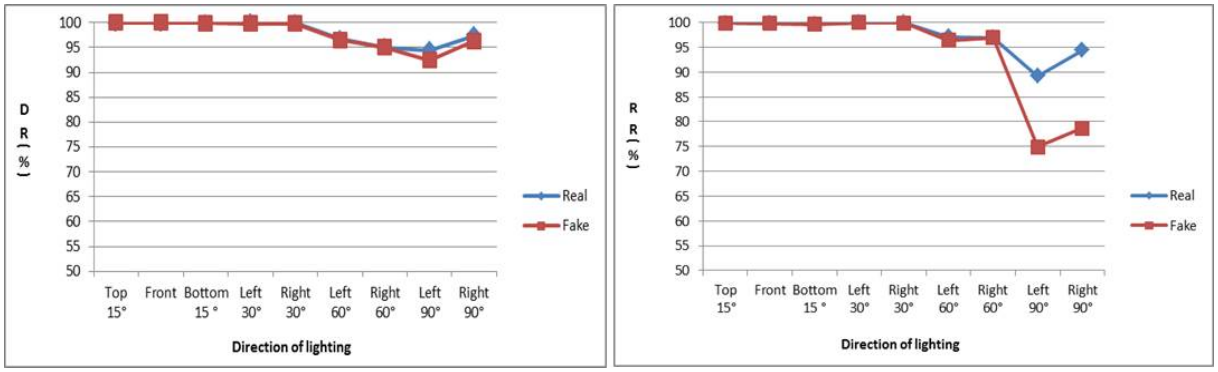
Each engine registered five real face images for each subject under normal lighting conditions. The test engine obtained recognition results from every frame in the real and fake facial videos; we calculated the recognition rate from the detected faces only.

Table 1: Performance comparison for each engine

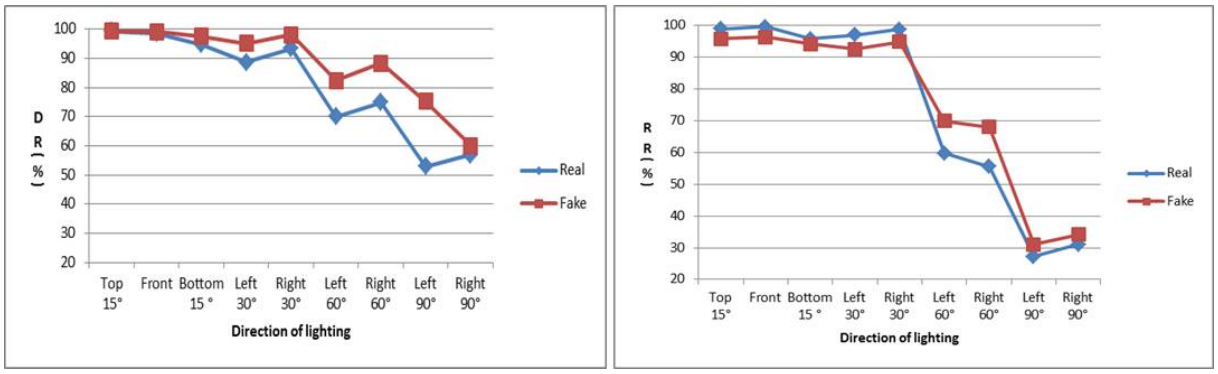
Engine	Detection rate(%)			Recognition rate(%)		
	Real faces	Fake faces	Deviations	Real faces	Fake faces	Deviations
A	98.15	97.77	0.38	97.47	94.02	3.45
B	80.86	88.48	-7.62	73.80	75.13	-1.33
C	51.69	46.65	5.04	58.79	53.57	5.22

Table 1 illustrates the detection and recognition rate deviations of the real facial images and fake facial videos for each engine. Even though each engine exhibited a different recognition performance, the deviations between the real facial images and fake facial videos were all less than 5.22%. In other words, there is no significant difference in facial recognition performance when using fake facial videos

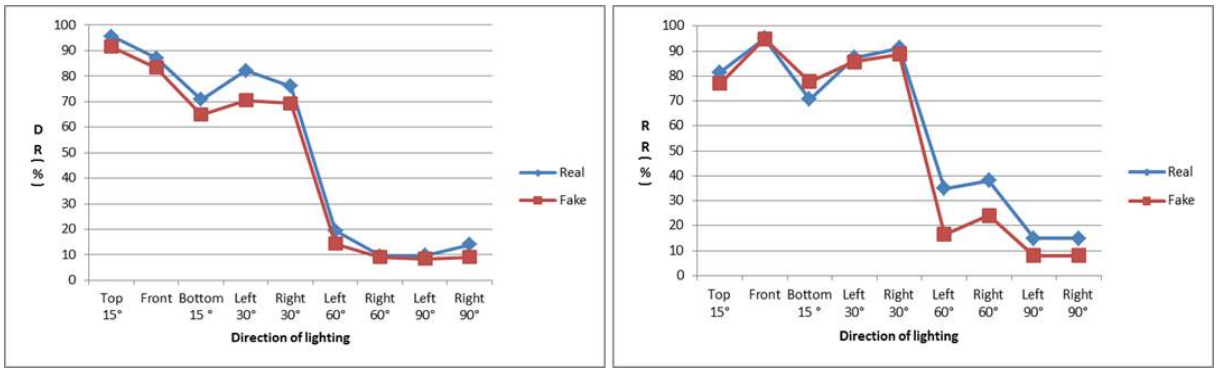
instead of real faces.



(a) Change in performance of engine A



(b) Change in performance of engine B



(c) Change in performance of engine C

Figure 6: Changes in performance according to lighting direction

Figure 6 shows performance changes according to the lighting direction for each engine. The y-axis represents recognition rate and detection rate. The x-axis represents lighting direction, with poor lighting conditions to the left. Typically, face recognition performance deteriorates under poor lighting conditions. The better the recognizer is, the wider the coverage range is.

In engine A, the performances of both real facial images and fake facial videos under left/right or 90° lighting were inferior to those under good lighting conditions. Performances of both engines B and C drastically declined under left/right or 60° lighting, because coverage was narrower than engine A's. Moreover, performance gaps between real faces and fake faces under poor lighting conditions increased. In conclusion, if a general web camera was equipped with the secure face recognition system, the system

would not be able to differentiate "live" faces from "not live" faces under good lighting conditions.

6 Conclusions

In this paper, we compared real faces with high-definition fake facial videos from LED display devices and showed face recognition performance changes according to lighting direction. With the development of high definition display devices, spoofing attacks using HD or UHD monitors will increase. Although poor lighting conditions have different effects depending on the engine, the secure face recognition system on general web cameras is not able to differentiate "live" faces from "not live" faces under good lighting conditions, like uniform lighting. In future works, we will expand on this experiment by using not only lighting direction, but lighting type and intensity of illumination.

Acknowledgments

This work is supported partly by the R&D program of the Korea Ministry of Trade, Industry and Energy(MOTIE) and the Korea Evaluation Institute of Industrial Technology (KEIT). (Project: Technology Development of service robot's performance and standardization for movement/manipulation/HRI/Networking, 10041834).

References

[1] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *Proc. of the 2009 International Conference on Image Analysis and Signal Processing (IASP'09), Taizhou, China*, pages 233–236. IEEE, April 2009.

[2] Saptarshi Chakraborty and Dhrubajyoti Das. An overview of face liveness detection. *International Journal on Information Theory*, 3(2):11–25, April 2014.

[3] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Proc. of the International Conference of the Biometrics Special Interest Group (BIOSIG'12), Darmstadt, Germany*, pages 1–7. IEEE, September 2012.

[4] CIE. International commission on illumination. `http://www.cie.co.at`, 2000.

[5] Center for Biometrics and Security Research. Casia face anti-spoofing database. `http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp`.

[6] Hyungkeun Jee, Sunguk Jung, and Janghee Yoo. Liveness detection for embedded face recognition system. *International Journal of Biomedical Sciences*, 1(4):235–238, October 2006.

[7] Gahyun Kim, Sungmin Eum, Jae Kyu Suhr, Dong Ik Kim, Kang Ryoung Park, and Jaihie Kim. Face liveness detection based on texture and frequency analyses. In *Proc. of the 5th IAPR International Conference on Biometrics (ICB'12), New Delhi, India*, pages 67–72. IEEE, March–April 2012.

[8] Sooyeon Kim, Sunjin Yu, Kwangtaek Kim, Yuseok Ban, and Sangyoun Lee. Face liveness detection using variable focusing. In *Proc. of the 6th 2013 International Conference on Biometrics (ICB'13), Madrid, Spain*, pages 1–6. IEEE, June 2013.

[9] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security*, 2(3):548–558, September 2007.

[10] Andrea Lagorio, Massimo Tistarelli, Marinella Cadoni, Clinton Fookes, and Sridha Sridharan. Liveness detection based on 3d face shape analysis. In *Proc. of the 2013 International Workshop on Biometrics and Forensics (IWBF'13), Lisbon, Portugal*, pages 1–4. IEEE, April 2013.

[11] Lin Sun, Gang Pan, Zhaohui Wu, and Shihong Lao. Blinking-based live face detection using conditional random fields. In *Proc. of the 2nd International Conference on Biometrics (ICB'07), Seoul, Korea, LNCS*, volume 4642, pages 252–260. Springer-Verlag, August 2007.

[12] Xiaoyang Tan. Nuaa photograph imposter database. `http://parnec.nuaa.edu.cn/xtan/data/nuaaimposterdb.html`, 2010.

[13] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Proc. of the 11th European Conference on Computer vision (ECCV'10), Heraklion, Greece, LNCS*, volume 6316, pages 504–517. Springer-Verlag, September 2010.

[14] Junjie Yan, Zhiwei Zhang, Zhen Lei, Dong Yi, and Stan Z. Li. Face liveness detection by exploring multiple scenic clues. In *Proc. of the 2013 International Conference on Control, Automation, Robotics and Vision (ICARCV'12), Guangzhou, China*, pages 188–193. IEEE, December 2012.

[15] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z. Li. Face liveness detection with component dependent descriptor. In *Proc. of the 6th 2013 International Conference on Biometrics (ICB'13), Madrid, Spain*, pages 1–6. IEEE, June 2013.

Author Biography

Mi-Young Cho is a senior researcher at the Electronics and Telecommunications Research Institute. She received her BS degrees in information and telecommunication engineering from Chosun University in 2002, and her MS and PhD degrees in computer science from Chosun University in 2004 and 2008, respectively. Since 2009, she has been researching the testing and evaluation of service robots at the Electronics and Telecommunications Research Institute.

Young-Sook Jeong is a principal researcher at Electronics and Telecommunications Research Institute. She received her BS degrees in computer science from the Ewha Womans University in 1988, and her MS degree in electronic engineering from the Chungnam National University in 2001, respectively. She has research experience in the field of the performance evaluation and standardization for service robot.