A Dataset and Benchmark for Large-scale Multi-modal Face Anti-spoofing -Supplementary Material-

Shifeng Zhang[1*], Xiaobo Wang[2*], Ajian Liu[3], Chenxu Zhao[2],
Jun Wan[1†], Sergio Escalera[4], Hailin Shi[2], Zezheng Wang[5], Stan Z. Li[1,3]
[1]NLPR, CASIA, UCAS, China; [2]JD AI Research; [3]MUST, Macau, China
[4]Universitat de Barcelona, Computer Vision Center, Spain; [5]JD Finance
{shifeng.zhang,jun.wan,szli}@nlpr.ia.ac.cn, ajianliu92@gmail.com
{wangxiaobo8,zhaochenxu1,shihailin,wangzezheng1}@jd.com, sergio@maia.ub.es

1. CASIA-SURF dataset samples

More samples of the proposed CASIA-SURF dataset are shown in Fig. 1. We select 6 subjects with their corresponding real (left) and fake (right) images. Each real or spoofing attacked sample contains 10 images with 3 modalities (*i.e.*, RGB, Depth and IR). All the 6 attack styles are shown in this figure. One can observe that:

- It is very challenging to distinguish between fake and real images in attack 1 and 2 just considering RGB images. In attack 1 and 2, only the eyes are cut from the face region. The fake face retains the texture very well.
- Our dataset includes subjects wearing eyeglasses. The eyeglass is easy to be analyzed for real *v.s* fake identification from Depth and/or IR images in attack 3 and 4.
- The depth image provides the most discriminative features for all six attack styles, especially for attack 1 and 2 where depth features of nose and mouth are missing.
- Different modalities provide the complementary information, and face anti-spoofing detection can benefit from multi-modal data fusion.

From Table 4 in the main manuscript, one can also see that when only a single modality is used for training, the best performance is achieved by depth information. Overall, the best performance is achieved when all three modalities are used.

2. Qualitative results

The quantitative results of different visual modalities have already been listed in Table 4 of the main manuscript with the ROC curve as the evaluation metric. Here, we present some false positive (FP) and false negative FN samples shown in Fig. 2. Labels from (1) to (7) are recognized results by the models trained with different combinations of modalities: RGB, Depth, IR, RGB&Depth, RGB&IR, Depth&IR and RGB&Depth&IR. It can be observed that these selected samples are challenging for face anti-spoofing detection, even many of them are very difficult to be distinguished by human eyes.

*These authors contributed equally to this work
†Corresponding author

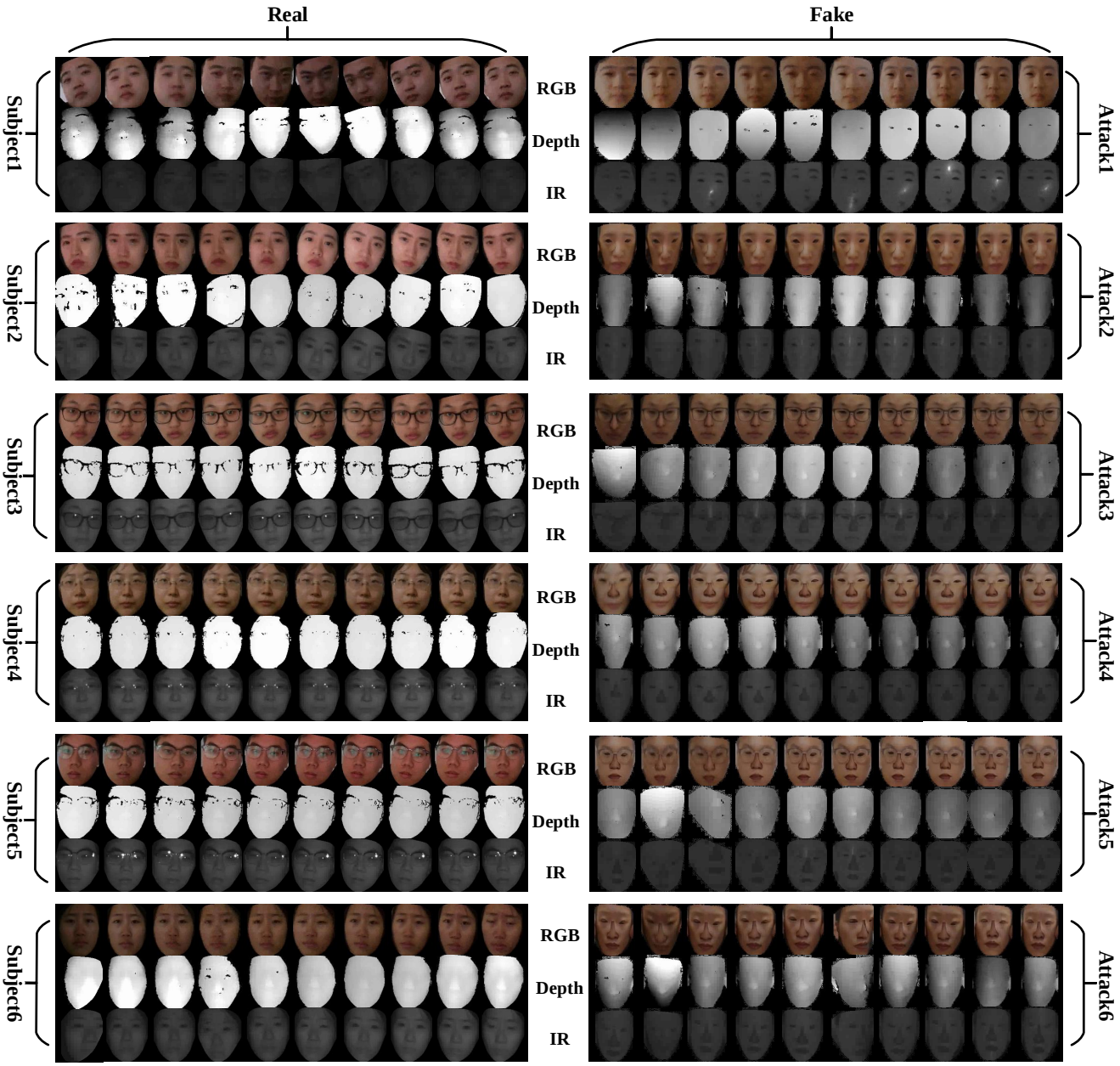

Figure 1. Samples of the CASIA-SURF dataset. Real and fake images are shown for six subjects and six attack types. Each sample has RGB, Depth and IR images.

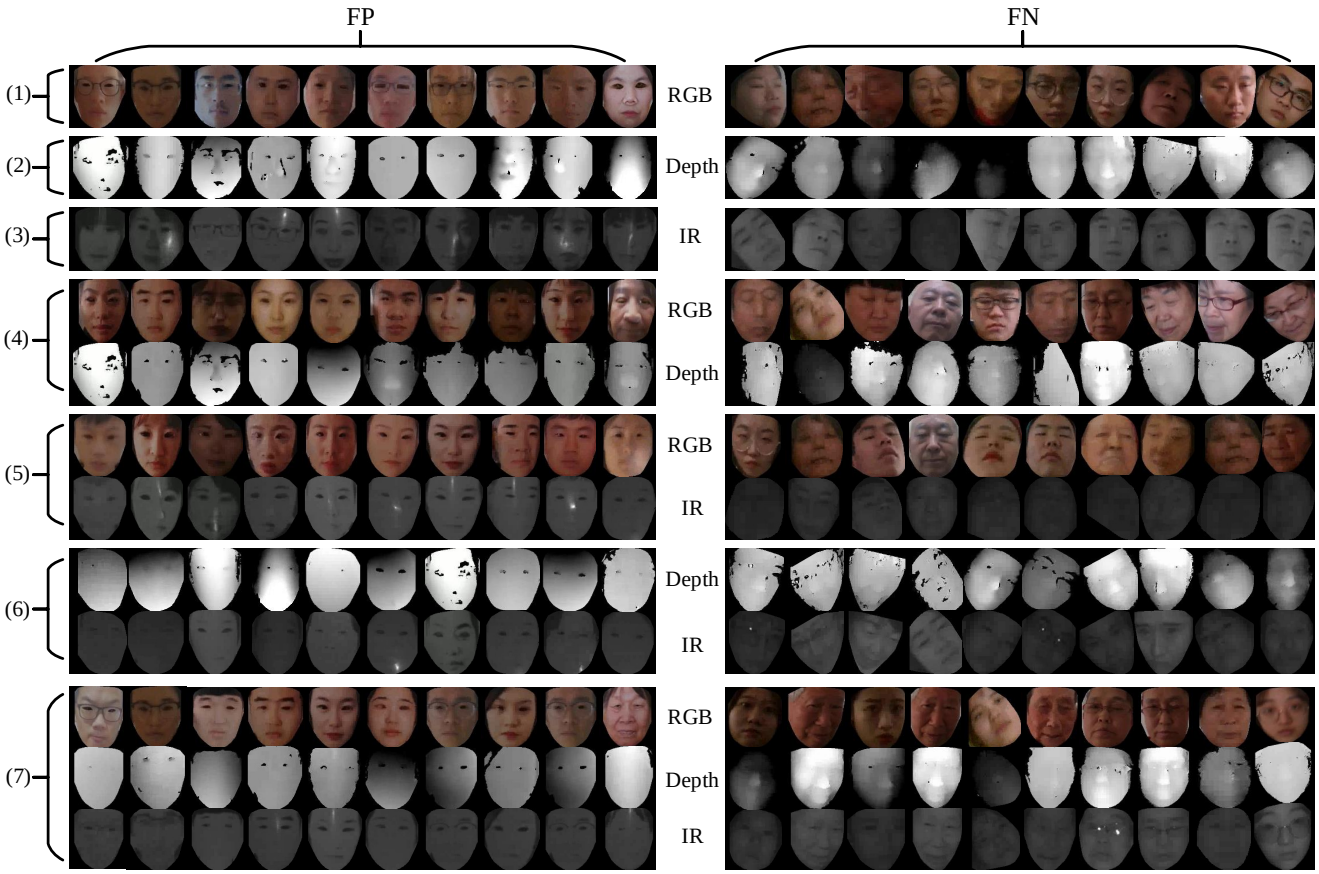

Figure 2. Difficult samples in the CASIA-SURF dataset. Left column: false positive samples mean the fake sample is wrongly recognized as the real one. Right column: false negative samples mean the real sample is wrongly recognized as the fake one. The labels from (1) to (7) correspond to the recognition results of models trained with combinations of different modalities: RGB, Depth, IR, RGB&Depth, RGB&IR, Depth&IR and RGB&Depth&IR.