OULU-NPU: A mobile face presentation attack database with real-world variations

Database description:

The OULU-NPU face presentation attack database consists of 4950 real access and attack videos that were recorded using front facing cameras of six different smartphones in the price range from €250 to €600. The real videos and attack materials were collected in three sessions with different acquisition conditions (Session 1, Session 2 and Session 3). In order to simulate realistic mobile authentication scenarios, the video length was limited to five seconds and the subjects were asked to hold the mobile device like they were being authenticated but without deviating too much from their natural posture while normal device usage. The attack types considered in the OULU-NPU database are print and video-replay. The attacks were created using two printers (Printer 1 and Printer 2) and two display devices (Display 1 and Display 2).

Files:

- The folder 'Train_files' contains 360 real and 1440 attack videos of 20 subjects. These files are used to train the countermeasure methods.
- The folder 'Dev_files' contains 270 real and 1080 attack videos of 15 subjects. These files are used to tune the countermeasure methods.
- The folder 'Test_files' contains 360 real and 1440 attack videos of 20 subjects. These files are used to evaluate the performances of the countermeasure methods.

In addition to the video files, we provide also the eye locations for each video as text files. The eye locations have been automatically determined using the dlib[1] library. The organization of these video files is as follows:

num_frame, x_eye_left, y_eye_left, x_eye_right, y_eye_right

- num*_frame=0* corresponds to the first frame.
- *x_eye_left=0, y_eye_left=0, x_eye_right=0, y_eye_right=0* depicts that the face or facial landmark detector has failed to detect the face or landmarks in the corresponding frame.

The naming convention for the video and the eye location files is:

Phone_Session_User_File.avi / Phone_Session_User_File.txt,

where:

- **Phone (1…6)** is the phone ID.
- **Session (1…3)** is the session number.
- **User (1…55)** is the subject ID. The users from 1 to 20 are used for training, from 21 to 35 for development and 36 to 55 for testing.
- **File (1…5)** is the access type, where: 1=real; 2=print1; 3=print2; 4=video-replay1; 5=video-replay2**.**

[1] http://dlib.net/imaging.html

Evaluation protocols:

For the evaluation of the generalization capability of the face PAD methods, four protocols are used.

Protocol I:

The first protocol is designed to evaluate the generalization of the face PAD methods under previously unseen environmental conditions, namely illumination and background scene. As the database is recorded in three sessions with different illumination condition and location, the train, development and evaluation sets are constructed using video recordings taken in different sessions.

Protocol II:

The second protocol is designed to evaluate the effect of attacks created with different printers or displays on the performance of the face PAD methods as they may suffer from new kinds of artifacts. The effect of attack variation is assessed by introducing a previously unseen print and video-replay attack in the test set.

Protocol III:

One of the critical issues in face PAD and image classification in general is sensor interoperability. To study the effect of the input camera variation, a Leave One Camera Out (LOCO) protocol is used. In each iteration, the real and the attack videos recorded with five smartphones are used to train and tune the algorithms, and the generalization of the models is assessed using the videos recorded with the remaining one.

Protocol IV:

In the last and most challenging protocol, all above three factors are considered simultaneously and generalization of face PAD methods are evaluated across previously unseen environmental conditions, attacks and input sensors.

The following table gives a detailed information about the video recordings used in the train, development and test sets of each protocol. **P** refers to printer and **D** refers to display.

Protocol	Subset	Session	Phones	Users	Attack created using:	Real videos	Attacks videos	All videos
Protocol I	Train	Session 1,2	6 Phones	1-20	P 1,2; D 1,2	240	960	1200
	Dev	Session 1,2	6 Phones	21-35	P 1,2; D 1,2	180	720	900
	Test	Session 3	6 Phones	36-55	P 1,2; D 1,2	120	180	600
Protocol II	Train	Session 1,2,3	6 Phones	1-20	P 1; D 1	360	720	1080
	Dev	Session 1,2,3	6 Phones	21-35	P 1; D 1	270	540	810
	Test	Session 1,2,3	6 Phones	36-55	P 2; D 2	360	720	1080
Protocol III	Train	Session 1,2,3	5 Phones	1-20	P 1,2; D 1,2	300	1200	1500
	Dev	Session 1,2,3	5 Phones	21-35	P 1,2; D 1,2	225	900	1125
	Test	Session 1,2,3	1 Phone	36-55	P 1,2; D 1,2	60	240	300
Protocol IV	Train	Session 1,2	5 Phones	1-20	P 1; D 1	200	400	600
	Dev	Session 1,2	5 Phones	21-35	P 1; D 1	150	300	450
	Test	Session 3	1 Phone	36-55	P 2; D 2	20	40	60

For each protocol, the files 'Train.txt', 'Dev.txt' and 'Test.txt' contain, respectively, the list of the video files that will be used for training, tuning and evaluating the proposed methods. These files are organized as follows:

+1, filename_1

-1, filename_2

…

+1, filename_i

+1: means the corresponding video file is real access.

-1: means the corresponding video file is presentation attack.

In the lists corresponding to the evaluation sets -1 means the corresponding video file is print attack and -2 means that the corresponding video file is replay attack.

Please note that in the protocol 3 and 4, the Leave One Camera Out scenario is used. Thus, there are six training and development subsets (Train_i.txt, Dev_i.txt, Test.txt: i=1…6). These subsets will be used to train, tune and evaluate six different models.

Baseline method

The Matlab source codes of a color texture analysis based face PAD method [1] is provided as a baseline. After choosing ten random frames from each video, the LBP features are extracted from 64x64 images in the YCbCr and HSV color spaces. The resulting histograms computed over the color spaces are concatenated and fed into a Softmax classifier. The ten face images are classified separately and the average of the resulting scores is used as a final score for the whole video sequence.

[1] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *IEEE International Conference in Image Processing (ICIP)*, Quebec City, 2015, pp. 2636-2640.

Acknowledgements

If you use this database, please cite the following publication:

```
@INPROCEEDINGS{OULU_NPU_2017,
      author = {Boulkenafet, Z. and Komulainen, J. and Li, Lei. and Feng, X. and Hadid, A.},
    keywords = {biometric, Counter-Measures, Local Binary Patterns, Spoofing Attacks},
       month = May,
       title = {{OULU-NPU}: A mobile face presentation attack database with real-world variations},
     journal = {IEEE International Conference on Automatic Face and Gesture Recognition},
        year = {2017},}
```