Face Anti Spoofing: Handcrafted and Learned Features for Face Liveness Detection

Paulo Costa[1]
1101413@isep.ipp.pt
Pedro Silva[2]
up201604470@fe.up.pt
João Ribeiro Pinto[2,3]
joao.t.pinto@inesctec.pt
Ana F. Sequeira[3]
ana.f.sequeira@inesctec.pt
Ana Rebelo[3,4]
ana.m.rebelo@inesctec.pt

[1] Instituto Superior de Engenharia do Porto
Porto, Portugal
[2] Faculdade de Engenharia da Universidade do Porto
Porto, Portugal
[3] INESC TEC
Porto, Portugal
[4] Universidade Portucalense
Porto, Portugal

Abstract

The development of presentation attack detection (PAD) methods has become a high level concern in biometric security. As in other pattern recognition tasks, the use of deep learning is increasingly common. However, it is still doubtful if handcrafted features should be discarded. This work focused on the comparison of using handcrafted features and deep learning techniques at the feature extraction level in a face PAD method. Handcrafted features were based on Local Binary Patterns, while a Convolutional Neural Network based on VGG-16 was used for deep feature extraction. A Support Vector Machine was used for binary classification after dimensionality reduction using Principal Component Analysis. The methods were tested using the NUAA database, and the results show that handcrafted feature extraction still offer better results, with 3.1% APCER and 25.2% BPCER.

1 Introduction

Biometric systems are currently used in several different application scenarios. Border control, military facilities, and the mobile access to personal accounts or banking operations are some of the applications that require high reliability and robustness levels. Face biometric recognition is currently one of the most common traits for several applications due to its advantages over other biometric traits.

As face biometric recognition is increasingly used for access control and authentication, guarding sensitive information and valuable goods, the motivation to attack such systems is growing. Face recognition systems can be attacked using printed photographs, masks, or video displays [1, 4, 8, 9]. Presentation attack detection (PAD) methods have been proposed to tackle this problem in face recognition systems. These feature extraction methods are classified in previous studies into two categories of non-training-based and training-based.

Among non-training-based methodologies, Tan *et al.* [9] used a sparse low-rank bilinear discriminative model on image features extracted by a difference of Gaussian (DoG) and/or logarithmic total variation (LTV) methods to discriminate the *bona fide* and presentation attack images. Using the NUAA database [9], they showed that the *bona fide* and presentation attack images can be discriminated using their proposed method.

With the NUAA database, Määtta *et al.* [4] used three feature extraction methods (Gabor filter, local phase quantization, and local binary patterns LBP) to extract the image features and classify the *bona fide* and presentation attack images using support vector machines (SVMs). The classification error was significantly reduced compared to those obtained by Tan et al. *et al.* [9]. Benlamoudi *et al.* [1] also used the LBP method for addressing the PAD problem for a face recognition system. They used the Fisher score to reduce the dimensionality of the extracted features.

Parveen *et al.* [8] proposed a method that uses a dynamic local ternary pattern (DLTP) for detecting presentation attack face images. They used the DLTP method to extract image features and the SVM method for classification. However, the detection accuracy when using the NUAA database was slightly worse than those obtained by Määtta *et al.* [4] and Benlamoudi *et al.* [1]. These results show that the handcrafted features are suitable for presentation attack detection. However, as verified in these studies, the detection accuracy varies significantly among different databases, indicating some of the handicaps of handcrafted features.

Recently, deep learning frameworks have frequently overcame conventional methods in tasks like image classification, object detection, or face-based age estimation. Considering this, training-based feature extraction methods have been studied by Menotti *et al.* [3]. The results indicate the sufficiency of the deep learning method for detecting presentation attack images in biometric recognition systems. However, the method was unable to outperform the handcrafted feature extraction method in all cases.

In a study by Nanni *et al.* [5], the deep learning framework was applied for general image classification problem as an image feature extractor. In detail, they used several CNN models which were trained for several different problems to extract image features of the current problem. Based on the extracted image features, they used several SVM models to classify the input images into desired classes.

In another study [6], the authors additionally used several kinds of handcrafted feature extraction methods such as LBP or local ternary patterns (LTP) alongside deep networks. The results show that the handcrafted and deep image features can extract different information from input images. Based on this result, they showed that the combination of handcrafted and deep features is sufficient to lead to an enhancement of classification accuracy.

Nguyen *et al.* [7] proposed a new PAD method based on hybrid features that combines information from both handcrafted and deep learning features. This is the first approach to PAD for face recognition systems using a combination of deep and handcrafted image features. By combining the deep and handcrafted image features, they enhanced the detection accuracy compared to conventional state-of-the-art detection methods and reduce the variation in detection accuracy caused by the variation in face images.

This work seeks to compare handcrafted and deep feature extraction, and their impact in presentation attack detection performance. It aims to lay the foundations for future work in the application of deep learning for biometric presentation attack prevention.

2 Proposed Methodology

Pipelines of the studied methods are depicted in Fig. 1. A face image is received and processed to produce a binary decision of *bona fide* or *presentation attack*. The first step of the method is to extract the image features using handcrafted algorithm local binary pattern (LBP) or convolutional neural network (CNN) or hybrid (LBP + CNN).

The second step is to apply principal component analysis (PCA) for dimensionality reduction. The models are tested with and without PCA to test which present better performance. At last, SVM is used to classify images as *bona fide* or as presentation attacks. Below, the feature extraction and classification processes are described in detail.

Feature extraction: The LBP feature extraction method has been used to extract image features for many computer vision tasks, including face recognition, with advantages in illumination and rotation invariance [2]. The LBP method computes a p-bit binary descriptor for each pixel in a given image using its surrounding pixels. For this study, the LBP used number of points $p = 8$ and radius $r = 1$. For the CNN, the VGG-16 architecture was used, with VGG-Face pretrained weights. As the dense layers were discarded, the outputs of the last pooling layer of the net-

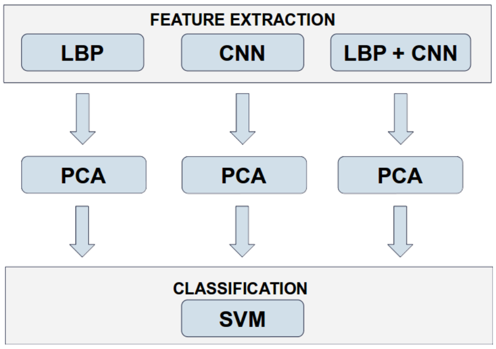

Figure 1: Schema of the proposed method.

Feature Extraction	SVM Kernel	APCER	BPCER
LBP	Sigmoid	11.005%	20.205%
CNN	RBF	37.121%	34.022%
LBP$_{w/PCA}$	**Linear**	**3.093%**	**25.221%**
CNN$_{w/PCA}$	Poly	20.375%	32.390%
(LBP+CNN)$_{w/PCA}$	Poly	19.393%	32.408%

Table 1: Summary of the main results of the implemented methodologies.

work were used as feature vectors. Hybrid LBP and CNN feature vectors were obtained through the concatenation of the individual feature vectors from each extraction process. Because the hybrid feature vector is a combination of two types of feature vectors, it should contain more useful information for presentation attack detection.

Classification: Support Vector Machines were used to find the largest-margin hyperplane that can separate the samples of each class. In this work, to allow for non-linearity, the SVM was explored with four different kernels (radial basis function, sigmoid, linear, and polynomial) with hyperparameters optimised through grid-search.

3 Experimental Setup

Data: The NUAA database is a public database for training and evaluating PAD methods for face recognition systems [9]. This database simulates a simple and general method that re-captures a printed photograph of users for attacking a face recognition system. It contains *bona fide* and presentation attack face images from 15 individuals. For each person, they captured both *bona fide* and presentation attack images in three different sessions using generic cheap webcams and *bona fide* face and printed subject photographs. The database contains 5105 *bona fide* and 7509 presentation attack face images in colour space with size 640×480. This work used the predefined training and testing subsets, to enable direct comparison with the literature. The training database contains 1743 *bona fide* and 1748 presentation attack face images, while the testing database contains 3362 *bona fide* and 5761 presentation attack face images.

Evaluation metrics: To evaluate the performance of the PAD methods, two metrics were used: the *attack* presentation classification error rate (APCER) and the *bona-fide* presentation classification error rate (BPCER). By definition, APCER indicates the proportion of *attack* presentations incorrectly classified as *bona fide* presentations, while BPCER indicates the proportion of *bona-fide* images incorrectly classified as *attack* images.

4 Results and discussion

The performance results are presented in Table 1. A detection error (APCER) of 11.01% was attained using LBP features and sigmoid SVM. However, using PCA, the detection errors are further reduced to 3.09 using linear kernel of the SVM method. These experimental results indicate that the PCA method is useful to reduce the dimensionality of the image features and to enhance the detection accuracy of the PAD method using handcrafted image features on the NUAA database.

Using features extracted using the VGG-Face network, with PCA for reduction, we obtain the smallest APCER with Polynomial SVM kernel, 20.38%. However, both errors are much higher than those attained using LBP features. The hybrid solution (LBP + CNN) improved the results of the experiments using only CNN extracted features. However, the sole use of LBP features offered better results than the hybrid version. Considering this, the VGG-Face architecture and weights are likely not adequate for the task of detecting presentation attacks. Moreover, the use of SVM limits the potential of deep learning methodologies. Hence, the use of an end-to-end CNN should be further explored.

5 Conclusions and future work

This work consisted on the study of different methods to detect presentation attack images for a face recognition system. The results indicate that LBP-based handcrafted features are more suitable to detect presentation attacks than VGG-Face features. Additionally, PCA enhanced the performance of the method. Nevertheless, further efforts should be devoted to the study of end-to-end deep networks for PAD, to improve performance and overcome the limitations of current handcrafted solutions.

Acknowledgements

This work was financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalization - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project “POCI-01-0145-FEDER-030707”, and within the PhD grant “SFRH/ BD/137720/2018”.

References

[1] A. Benlamoudi, D. Samai, A. Ouafi, S. E. Bekhouche, A. Taleb-Ahmed, and A. Hadid. Face spoofing detection using local binary patterns and fisher score. In *CEIT 2015*, 2015. doi: 10.1109/CEIT.2015.7233145.

[2] W. O. Lee, Y. G. Kim, H. G. Hong, and K. R. Park. Face Recognition System for Set-Top Box-Based Intelligent TV. *Sensors*, 14(11): 21726–21749, 2014. doi: 10.3390/s141121726.

[3] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):864–879, 2015. doi: 10.1109/TIFS.2015.2398817.

[4] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *IJCB 2011*, 2011. doi: 10.1109/IJCB.2011.6117510.

[5] L. Nanni and S. Ghidoni. How could a subcellular image, or a painting by Van Gogh, be similar to a great white shark or to a pizza? *Pattern Recognition Letters*, 85, 2017. doi: 10.1016/j.patrec.2016.11.011.

[6] L. Nanni, S. Ghidoni, and S. Brahnam. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172, 2017. ISSN 0031-3203. doi: 10.1016/j.patcog.2017.05.025.

[7] D. T. Nguyen, T. D. Pham, N. R. Baek, and K. R. Park. Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors. *Sensors*, 18(3), 2018. doi: 10.3390/s18030699.

[8] S. Parveen, S. M. S. Ahmad, N. H. Abbas, W. A. W. Adnan, M. Hanafi, and N. Naeem. Face Liveness Detection Using Dynamic Local Ternary Pattern (DLTP). *Computers*, 5(2), 2016. doi: 10.3390/computers5020010.

[9] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *ECCV 2010*, pages 504–517, 2010.