2017 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 25–28, 2017, TOKYO, JAPAN

IMPROVING IMAGE CLASSIFICATION WITH FREQUENCY DOMAIN LAYERS FOR FEATURE EXTRACTION

José A. Stuchi[1,2], *Marcus A. Angeloni*[2], *Rodrigo F. Pereira*[2], *Levy Boccato*[2], *Guilherme Folego*[2,3], *Paulo V. S. Prado*[1,4], *Romis R. F. Attux*[2]

[1] Phelcom Technologies, São Carlos [2]State University of Campinas, Campinas
[3]CPqD Foundation, Campinas [4]University of São Paulo, São Carlos
São Paulo State, Brazil

ABSTRACT

Machine learning has been increasingly used in current days. Great improvements, especially in deep neural networks, helped to boost the achievable performance in computer vision and signal processing applications. Although different techniques were applied for deep architectures, the frequency domain has not been thoroughly explored in this field. In this context, this paper presents a new method for extracting discriminative features according to the Fourier analysis. The proposed frequency extractor layer can be combined with deep architectures in order to improve image classification. Computational experiments were performed on face liveness detection problem, yielding better results than those presented in the literature for the *grandtest* protocol of Replay-Attack Database. This paper also aims to raise the discussion on how frequency domain layers can be used in deep architectures to further improve the network performance.

Index Terms— machine learning, Fourier analysis, image classification, liveness detection, deep learning

1. INTRODUCTION AND RELATED WORK

Artificial neural networks and deep learning have revolutionized many applications in the last years, especially in computer vision and signal processing areas [1].

In simple terms, an artificial neural network is a machine designed according to a model of how the human brain processes a particular activity or a function of interest. The brain has a plasticity that enables the nervous system to adapt itself according to the environment in which it is inserted. Artificial neural networks are composed of artificial neurons that have an adaptive structure (or plasticity) which allows them to learn from presented data [2]. In order to reach a good performance, neural networks use massive connections of artificial neurons. Based on the adopted neuron model and on the network architecture, the system has the capacity to learn and store experimental knowledge, and to make it available to further applications [2].

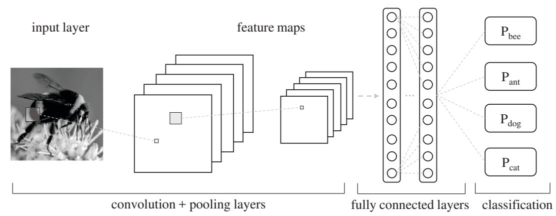


Fig. 1. Basic ConvNet architecture, adapted from [3].

An important advance in the field of neural networks, deep learning has been frequently studied in the literature due to the excellent results obtained in solving complex problems, being the state-of-the-art in many areas, including speech processing and computer vision [3]. Several factors, such as the progress in learning algorithms, the availability of large training datasets and the possibility of using highly parallel and distributed hardware, such as graphic processing units (GPUs), certainly contributed to the success of deep neural networks.

In this scenario, many different deep architectures have been proposed, such as deep belief networks (DBN), auto encoders, long-short term memory (LSTM) and convolutional neural networks (CNNs or ConvNets) [1]. In the fields of image processing and computer vision, ConvNets stand out due to the excellent results, mainly in object detection and recognition problems.

ConvNets are based on the structure of the brain visual cortex, which is responsible for processing visual information. Each individual layer performs a very simple operation, such as a weighted linear sum of inputs or nonlinear functions such as threshold or competitive normalization [1]. Fig. 1 presents a basic ConvNet architecture.

One of the main advantages of ConvNets is that they perform well in processing raw data, *i.e.*, as they are captured. Thus, in computer vision problems, for example, the network

We thank FAPESP (2016/00985-5) for the financial support.

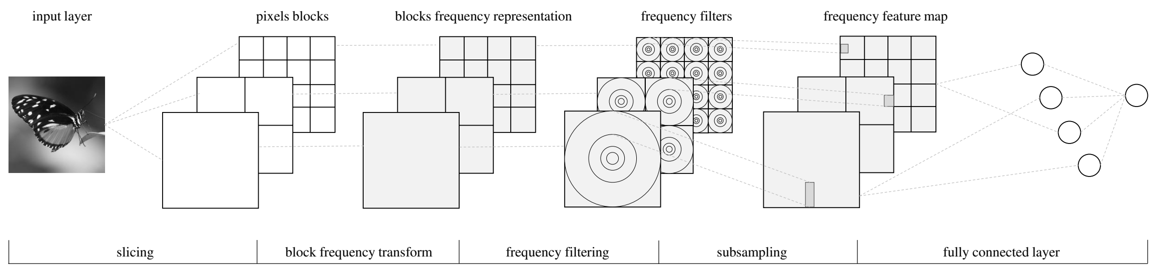


Fig. 2. Basic frequency layer architecture for image classification.

can learn directly from the image channels and, consequently, it does not need a high domain knowledge of a human specialist to manually design a feature extractor [2].

ConvNets employ essentially four main concepts: local connections, shared weights, grouping or pooling and the use of many layers [3]. These concepts can be seen in the different layers displayed in Fig. 1, which classify an image, generating at the output the probabilities for each class.

Based on the state-of-the-art results achieved in many different applications in computer vision, ConvNets are now being applied to several commercial solutions such as autonomous vehicles, object detection and recognition or even image labeling, when combined with LSTMs [3].

The present work faces ConvNets from a different perspective: converting data from the spatial domain to the frequency domain. Since ConvNets are based on convolution operations, and given that a convolution can be computed in the frequency domain through a multiplication [4], a new investigation branch can be opened to analyze and understand how features extracted by a network in the frequency domain can be useful to deep architectures. The appeal to use such technique is that some properties from the input image can be more easily detected on its spectrum (*e.g., Fourier Transform*) rather than on its raw pixels. While high frequency components represent both fine details (*e.g., soft borders*) and noise, low frequency components encode large scale information, such as shape and homogeneous areas [4].

Fourier analysis is widely used together with neural networks in signal processing applications such as brain EEG processing [5] and seismic analysis [6]. Applications in image processing are also explored, such as gesture recognition [7] and face detection [8].

However, few papers explore the frequency domain for image processing in deep learning field [9, 10]. In this context, this paper presents a feature extractor method based on frequency domain layers and raise the discussion on how these layers can be employed within deep architectures to improve the performance in computer vision problems.

Although the proposed method is generic and can be applied to different problems in computer vision, in this paper, it will be evaluated within the face biometrics scenario, more specifically for liveness detection, in order to avoid illegitimate access or spoofing attacks.

In a real authentication procedure, just one signal sampling process takes places, *i.e.*, the authentication device captures the image from the user's face. On the other hand, an attack to a face biometrics system might be performed by presenting a photograph, either on a printed paper or on a screen. In these attack cases, the sampling process happens more than once: thrice on a printed paper (first, a photograph is taken, then it is printed and finally acquired by the recognition device), or twice on a screen (photograph is taken then presented to the recognition device).

The re-sampling process degrades the output signal, given that the input signal (continuous) is quantized, giving rise to a discrete set of values. The output image is then blurred and fine details may be lost, which can be perceived in the frequency domain by the attenuation of high frequency components and the enhancement of low frequency components. Thus, frequency features can be useful for discriminating between authentic and fraudulent accesses.

Different approaches are presented in the literature for the face liveness detection problem using frequency domain information [11, 12, 13]. However, these methods usually apply fixed frequency bands, together with fixed thresholds, and do not analyze different image blocks, from the complete image to small textures.

Thus, this paper presents a new method to extract representative features from Fourier spectra by decomposing an image in small blocks in order to extract both global and local information, with different filtering bands in frequency. These bands are fixed in this preliminary work, inspired in the human eye response, but the idea is to train and update the bands in future works, analogously to convolutional kernels in ConvNets.

2. PROPOSED METHOD

The main idea behind the proposed method is to extract global and local discriminative features from the image using Frequency Domain Layers (FDL). Fig. 2 exhibits the basic idea underlying this method.

The raw image is the input to the pipeline flow. In order to extract global and local features, the image is sliced in four non-overlapping blocks in the first step. Each block is then sliced in four other blocks and so on. The idea is that big blocks will be able to represent global image features, such as background or an object main components, while small blocks represent local features, such as textures and corners. Only three slice levels were presented for the sake of clarity, but this can be modified according to the problem at hand. Thus, the number of blocks (or slice levels) and their sizes are parameters of the proposed method.

In the second step, the 2D Discrete Fourier Transform (DFT) [4] is applied to each block, from global to local regions. For each image block i, denoted $f_i(x, y)$ and with size $M_i \times N_i$, a 2D DFT is applied, defined as:

$$F_i(u, v) = \sum_{x=0}^{M_i-1} \sum_{y=0}^{N_i-1} f_i(x, y) e^{-j2\pi(ux/M_i + vy/N_i)} \quad (1)$$

Next, using each block magnitude spectrum, frequency filtering can be applied: in the third step, different filter bands are used to filter specific frequency ranges in the image blocks. Even though it would be possible to train and adjust the band cutoff frequencies (*e.g.* using the backpropagation algorithm), as done for updating convolutional kernels in ConvNets [1], they remain fixed in this first analysis.

The inspiration to determine cutoff frequencies for each band comes from the biological human eye response, which resolves low frequencies better than high frequencies, according to the model proposed in [14]. Based on this model, the eye Modulation Transfer Function (MTF) can be approximated by a negative exponential curve, meaning that the eye response quickly decays when frequency increases. Thus, the idea adopted in this paper is to use different band sizes for filtering, exponentially increasing from low to high frequencies. A similar idea was already applied in sound processing, as in the extraction of Mel-frequency cepstral coefficients [15], which was inspired in the human auditory system.

Fig. 3 presents six frequency bandpass filters. Since zero-frequency components were shifted to the center of spectrum, low frequencies are close to the square center and high frequencies are close to the border. In this example, the bands increase exponentially as frequency grows. Each band $B_k(u, v)$ is defined as:

$$B_k(u, v) = \begin{cases} 1, & \forall u, v \mid 2^{k-1} \leq \sqrt{u^2 + v^2} < 2^k \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where k is the band index. In the example of Fig. 3, $k \in [1, 6]$ and the image size is 128×128. An ideal binary filter was used, but other approaches can also be applied, such as Gaussian and Butterworth filters [4]. Using these bands, the filtering step can be applied. In this case, an element-wise matrix multiplication will be used, as follows:

$$R_{i,k}(u, v) = F_i(u, v) \odot B_k(u, v). \quad (3)$$

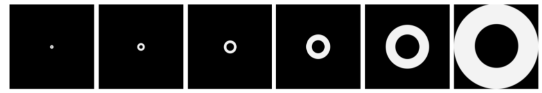


Fig. 3. Example of frequency bands filters, from low (left) to high (right) frequencies.

In the example displayed in Fig. 3, 6 filtered spectra will be obtained for the first image block, *i.e.*, from $R_{1,1}$ to $R_{1,6}$. In the smaller image blocks, resulting from slicing steps, filtering will be similarly performed, but now reducing the number of used bands. For the first slicing, the number of bands will be 5, whereas, for the second slicing, it will be 4. This happens since the image size is halved in each slicing, *i.e.*, it is reduced to 64×64 for the first slicing and 32×32 for second. Thus, $k \in [1, 5]$ for the blocks resultant of the first slicing, $k \in [1, 4]$ for the blocks of second slicing and so on.

After each spectrum is filtered, the subsampling (fourth) step can now be applied. In this case, different strategies can be used such as averaging or max pooling [1], or even extracting some statistical moments for each band, (*e.g.* mean, variance, skewness). This stage reduces the data dimensionality and, as a result, a frequency feature map will be obtained, which can be used in a pattern recognition application.

Finally, the extracted feature map serves as the input of a fully connected network, in the fifth step, which will be trained to perform a specific task, such as image classification. Other classifiers could also be used in the last layer, such as Support Vector Machines (SVMs) or probabilistic classifiers [2].

It is important to remark that color images can also be processed since the proposed method can be applied for each color channel individually.

3. EXPERIMENTS

In order to evaluate the potential of the proposed method, the face anti-spoofing task was chosen, which is a prominent topic in biometric recognition. In this scenario, images from the human face are used and must be classified between two different classes: authentic or attack images.

Since the input data are videos, frames were first extracted from each one. Then, each frame was converted into grayscale and resized to a fixed default dimension, so that images had a standard reference, allowing an efficient face location. Afterwards, face detection was performed on the scaled image, applying the traditional method of Viola & Jones [16], which returns a square that delimits the face location.

With this location at hand, the face region was cropped, and a log filter was applied aiming to minimize illumination variations. After that, the filtered image was resized again, this time to a power of 2 close to the image height and width. For the used dataset, which will be described in the next subsection, all face images were resized to the default dimension of 128×128 pixels, so that all preprocessed images had the same standard size.

Then, with the extracted faces from the original images, the proposed method was applied. It started slicing the original face image of size 128×128 in 4 blocks of 64×64 pixels and 16 blocks of 32×32 pixels. For each block, the DFT was computed. These frequency blocks were then filtered in the third step, using 6 bands for the largest block, 5 bands for medium blocks, and 4 bands for the smallest blocks. Sub-sampling was then applied in the fourth step, calculating the first three statistical moments for each filtered spectrum. Thus, this step generated $1 \times 6 \times 3 = 18$ features for the biggest block, $4 \times 5 \times 3 = 60$ features for the medium ones, and $16 \times 4 \times 3 = 192$ features for the smallest blocks. Finally, in the fifth step, all the extracted features, viz., the 270 statistical moments, were normalized and then provided as inputs to a discriminative classifier. At this stage, two normalization procedures were considered – *min-max* and *standard* – which are defined, respectively, as follows:

$$\hat{\mathbf{x}} = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \tag{4}$$

$$\hat{\mathbf{x}} = \frac{\mathbf{x} - \text{mean}(\mathbf{x})}{\text{std}(\mathbf{x})}, \tag{5}$$

where $\mathbf{x}$ is each specific feature array in the feature map.

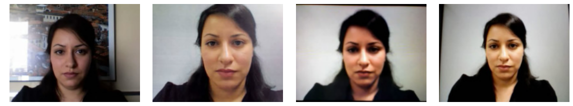
3.1. Dataset

The Replay-Attack Database [17] is a publicly available dataset that consists of $1{,}300$ short video recordings of both real-access and attack attempts of 50 different subjects. The movies were recorded on a laptop using its built-in webcam, producing color videos with 320×240 pixels, which were acquired under controlled and adverse lighting conditions. Face spoof attacks were generated by forging live verification attempts of the same subjects via printed photos, and images and videos displayed in a smartphone and also in a high resolution screen. Some samples of the real-access and attack attempts of this database are presented in Fig. 4.

3.2. Evaluation Methodology

Among the available protocols of Replay-Attack Database, the *grandtest* was chosen, which is the most challenging one, since it includes all types of devices to generate the attacks. In this protocol, the dataset is divided into three non-overlapping partitions: a training set, to optimize the algorithms; a development set, for threshold estimation; and a test set, where the final results are computed. Traditional measures used in biometrics were adopted to evaluate the algorithms: False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER), and Half Total Error Rate (HTER).

The False Acceptance Rate (FAR) is defined as the fraction of spoofing attacks improperly accepted by the system.



(a) (b) (c) (d)

Fig. 4. Samples from Replay-Attack Database: (a) real-access, (b) printed photo, (c) video displayed on mobile phone, and (d) photo displayed on high resolution screen.

Conversely, the False Rejection Rate (FRR) is the fraction of genuine access attempts incorrectly rejected by the system. The Equal Error Rate (EER) is used to summarize the performance of a system, defined as the error rate at a threshold point where $\texttt{FAR} = \texttt{FRR}$. Finally, the Half Total Error Rate (HTER) combines FAR and FRR at a specific threshold point, defined as $\texttt{HTER} = (\texttt{FAR} + \texttt{FRR})/2$.

3.3. Baselines

For comparison purposes, three types of state-of-the-art face anti-spoofing features from the client-independent discriminative baselines publicly available in [18] were considered. For all of them, SVM was used as discriminative classifier, with RBF kernel, decision surface regularization of $C = 1$, and kernel parameter controlling the support vectors influence $\gamma = 0$, which were the best values found in [18].

The first feature extractor was the Local Binary Patterns (LBP) [17], with the simplest uniform variant $LBP^{u2}_{8,1}$. The second aims at exploiting the variation of texture patterns in a spatial-temporal way, based on the LBP-TOP operator [19], also with the simplest regular uniform $LBP^{u2}_{8,8,8,1,1,1}$. Both were extracted from the face bounding box, normalized to 64×64, and a *min-max* normalization before the SVM training. The third baseline, MOTION [20], measures the correlation between the motion of the face and its background, with a *standard* normalization before the SVM training.

4. RESULTS AND DISCUSSIONS

Aiming to establish a fair comparison and to remove factors other than the proposed features, the same implementation of features normalization, classifier training parameters, scores generation and error rates computation was used for the baselines, available at [18]. The client-independent anti-spoofing approach was focused because it can be used in more operations of a biometric system (including enrollment), requiring only one classifier model for the whole system, and it can be integrated into the process of any biometric system already in use, without requiring the identity of clients. Table 1 shows the performance of baselines and the proposed method (FDL), using different slicing levels, considering the *grandtest* protocol of Replay-Attack Database.

Table 1. Comparison of different methods on *grandtest* protocol of Replay-Attack Database. Error rates are in percentage. *MM* stands for *min-max*, and *Std* for *standard*.

Method	**Norm**	**Dev**	**Test**		
		EER	**FAR**	**FRR**	**HTER**
LBP [17]	*MM*	14.56	9.56	21.29	15.42
LBP-TOP [19]	*MM*	8.19	4.45	12.60	8.53
MOTION [20]	*Std*	10.73	12.95	10.12	11.53
FDL-18	*Std*	14.24	12.84	10.92	11.88
FDL-18	*MM*	13.21	9.49	13.92	11.70
FDL-78	*Std*	14.70	**3.52**	12.32	7.92
FDL-78	*MM*	11.03	3.71	8.91	6.31
FDL-270	*Std*	9.09	9.96	**1.72**	5.84
FDL-270	*MM*	**5.92**	7.23	3.18	**5.21**

In the case of FDL-18, 18 frequency features are obtained directly from the 128×128 pixels original face image. On the other hand, FDL-78 considers one slicing level and extracts 60 features from 4 blocks of 64×64 pixels, which are concatenated with the previous 18 extracted features. Finally, FDL-270 involves two slicing levels, extracting 192 features from 16 blocks of 32×32 pixels, which are combined with the previous set of 72 features, generating an array of size 270.

It is possible to notice that Frequency Domain Layers outperformed the baselines in terms of HTER in the test set, achieving the best result with FDL-270 and *min-max* normalization. When no slicing was applied (FDL-18), the result was not satisfactory when compared to the baselines. However, performing slicing and concatenating the features from different blocks sizes proved to be pertinent in order to decrease the HTER value. In this case, it is possible to verify that features extracted from large blocks (global) and from small blocks (local) are complementary and, when used together, contribute for a performance improvement in the classification task. Additionally, the use of frequency bands of different sizes, being more granular on low frequencies and less granular in high frequencies, was useful for the system to achieve good results. Finally, the best HTER performance was attained when 270 features were considered and the *min-max* normalization procedure was applied. However, it is important to note that the best values of FAR and FRR were not achieved in this condition. In fact, the best FAR was achieved using only 78 features, demonstrating that features obtained from first slicing blocks were very discriminative. On the other hand, the best FRR was achieved using 270 features and *standard* normalization. Nevertheless, *min-max* achieved the best overall results, offering a better balance between rejecting spoofing attacks and accepting genuine authentications.

Regarding computational cost, the proposed method may enable real-time applications since it took about 4 ms to extract 270 features from a image of 128×128 pixels. Tests were performed using a Dell Optiplex 990 computer with Intel i7 3.40 GHz processor, 8 GB of RAM memory, Windows 7 Professional and OpenCV 3.1 library.

Therefore, it was observed that the proposed method was effective on face-liveness detection scenario, being a useful strategy to be applied in a standalone way or even in combination with other classifiers.

5. CONCLUSIONS

Deep learning has achieved excellent results in the last years for many machine learning problems. However, current deep architectures rarely exploit the frequency domain information for image processing. In this sense, this work presented a new method for extracting features from the Fourier frequency domain. The method applies the idea of iteratively dividing the image into smaller blocks in order to explore both global and local information. Additionally, it uses different sizes for the frequency bands used in the filtering step, inspired on the human eye response.

The method was tested on the face liveness detection context, more specifically, in the *grandtest* protocol of the Replay-Attack Database, which is the most challenging scenario. The obtained results present an improvement over the performances reported in the literature, reaching 5.21% HTER when using a vector with 270 frequency features for each frame. The obtained results, thus, were very satisfactory for the face-spoofing detection and motivate further tests in the future for new datasets and computer vision tasks.

The main contributions of this work correspond to the proposed feature extraction method in frequency domain, combining information from different block sizes and the use of different bands for filtering, being more granular in low frequencies than in high ones. These steps resulted in discriminative features that can be used in image classification problems. Nevertheless, there are some directions for future work that may lead to better results.

The first idea is to combine the proposed method with other classifiers. For instance, by combining the outputs of FDL-270 with LBP-TOP, the results may improve even more. The same idea can be used with deep architectures, *e.g.*, by employing the proposed feature extractor in association with a trained ConvNet.

The second idea is to train the cutoff frequencies for the frequency bands, as done for convolutional kernels in ConvNets, instead of using fixed bands as proposed. This may improve results, since bands will be placed in specific regions that are more convenient (and discriminative) to the problem at hand. For instance, specific band cutoff frequencies can be applied to classify textures in aerial images while other band frequencies can be employed to recognize faces. In this case, the frequency modeling will be more flexible, which may improve performance.

Increasing the number of slicing stages is another idea for future analysis. Two slicing levels were evaluated in this paper and the results improved as the image was divided in smaller blocks. Using even smaller blocks may bring even better results. In this case, a trade-off between the quantity of features and the classifier performance may be considered, since extracting more features requires a higher computational cost.

Finally, this work raises the discussion about how the proposed frequency domain layers can be useful when inserted in deep learning architectures. Combining different layers, in both space and frequency domain, in a deep network may improve the overall performance and it will be carefully investigated in future works.

6. REFERENCES

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT Press, 2016.

[2] Simon S Haykin, *Neural networks and learning machines*, Pearson Upper Saddle River, NJ, USA:, 3rd edition, 2009.

[3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] Rafael C. Gonzalez, Richard E. Woods, and Steven L. Eddins, *Digital Image Processing Using MATLAB*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

[5] Vairavan Srinivasan, Chikkannan Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, 2005.

[6] Farid U Dowla, Steven R Taylor, and Russell W Anderson, "Seismic discrimination with artificial neural networks: preliminary results with regional spectral data," *Bulletin of the Seismological Society of America*, vol. 80, no. 5, pp. 1346–1373, 1990.

[7] Peter RG Harding and Timothy Ellis, "Recognizing hand gesture using fourier descriptors," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. IEEE, 2004, vol. 3, pp. 286–289.

[8] Hazem M El-Bakry, "Face detection using fast neural networks and image decomposition," *Neurocomputing*, vol. 48, no. 1, pp. 1039–1046, 2002.

[9] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[10] Tom Brosch and Roger Tam, "Efficient training of convolutional deep belief networks in the frequency domain for application to high-resolution 2d and 3d images," *Neural computation*, 2014.

[11] Dhrubajyoti Das and Saptarshi Chakraborty, "Face liveness detection based on frequency and micro-texture analysis," in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on*. IEEE, 2014, pp. 1–4.

[12] 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 *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, 2012, pp. 67–72.

[13] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K Jain, "Live face detection based on the analysis of fourier spectra," in *Defense and Security*. International Society for Optics and Photonics, 2004, pp. 296–303.

[14] Isabel Escudero-Sanz and Rafael Navarro, "Off-axis aberrations of a wide-angle schematic eye model," *JOSA A*, vol. 16, no. 8, pp. 1881–1891, 1999.

[15] Steven Davis and Paul Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

[16] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I–511–I–518 vol.1.

[17] Ivana Chingovska, André Anjos, and Sébastien Marcel, "On the Effectiveness of Local Binary Patterns in Face Anti-spoofing," in *IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sept. 2012.

[18] Ivana Chingovska and André Rabello dos Anjos, "On the Use of Client Identity Information for Face Antispoofing," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 787–796, 2015.

[19] Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, pp. 1–15, 2014.

[20] André Anjos and Sébastien Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–7.