Face Spoofing Detection through Partial Least Squares and Low-Level Descriptors

William Robson Schwartz   Anderson Rocha   Helio Pedrini

Institute of Computing
University of Campinas, Campinas, SP, Brazil
Av. Albert Einstein, 1251, 13083-852
{schwartz,anderson.rocha,helio}@ic.unicamp.br

Abstract

Personal identity verification based on biometrics has received increasing attention since it allows reliable authentication through intrinsic characteristics, such as face, voice, iris, fingerprint, and gait. Particularly, face recognition techniques have been used in a number of applications, such as security surveillance, access control, crime solving, law enforcement, among others. To strengthen the results of verification, biometric systems must be robust against spoofing attempts with photographs or videos, which are two common ways of bypassing a face recognition system. In this paper, we describe an anti-spoofing solution based on a set of low-level feature descriptors capable of distinguishing between 'live' and 'spoof' images and videos. The proposed method explores both spatial and temporal information to learn distinctive characteristics between the two classes. Experiments conducted to validate our solution with datasets containing images and videos show results comparable to state-of-the-art approaches.

1. Introduction

Nowadays we are experiencing an increasing demand for highly secure identification and personal verification technologies. This demand becomes even more apparent as we become aware of new security breaches and transaction frauds [18]. In this context, biometrics has played a key role in the last decade providing tools and solutions either to verify or recognize the identity of a person based on physiological or behavioral characteristics. Among the used features are face, fingerprints, hand geometry, handwriting, iris, retinal vein, and voice.

Such methods, however, sometimes can be fooled (spoofed) by an identity thief, specially the ones based on face recognition, in which the thief can obtain a photo of an authentic user from a significant distance, or even obtain it from the Internet [17].

In some cases, a 2-D image-based facial verification or recognition system can be spoofed with no difficulty. As an example, consider the case in which one person, instead of showing his/her own face to a biometric system, displays a photo of an authorized counterpart either printed on a piece paper, on a laptop, or even on a cell phone screen. As a matter of fact, it is not unusual in practice to find some poorly-designed systems which have been shown to be fooled by very crude line drawings of a human face [13]. Although there has been important advances with respect to spoofing detection in the last decade, this research branch is still an open problem.

According to [15], there are three ways to spoof face recognition: (1) with a photograph of a valid user; (2) with a video of a valid user; or (3) with a 3D model of a valid user. In this paper, we focus on the attacks taking place considering the cases (1) and (2) above. These attacks can portray a printed face, a video sequence of a picture, or a video sequence of a dynamic scene (e.g., person moving the head or blinking). In all cases, the media used can be either moving (e.g., with a person holding the picture and slightly moving it) or fixed (e.g., picture attached to a tripod).

We can categorize current anti-spoofing methods into four non-disjoint groups: data-driven characterization, user behavior modeling, user interaction need, and presence of additional devices. Given the above categorization, non-intrusive methods without extra devices and human involvement may be preferable in practice, given that they could be integrated into an existing face recognition system, where usually only a generic webcam is deployed [17, 24]. Section 2 discusses some of the current anti-spoofing solutions for face-based recognition systems.

In this paper, we present an anti-spoofing solution based on a holistic representation of the face region, through a ro-

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

bust set of low-level feature descriptors, able to capture the differences between 'live' and 'spoof' images. Our solution explores both spatial and temporal information to learn differences between the two classes.

In order to validate the proposed solution, we perform tests with image- and video-based data sets. In tests with images, we compare our approach to the recent work of Tan et al. [24] over the public available NUAA data set. In the test with videos, we compare our approach to the research teams' results (including ours) that participated of the *2011 IJCB Competition on Counter Measures to 2D Facial Spoofing Attacks* [2] (referred to as FSA dataset in this work), using the videos and protocols of the competition.

The remainder of this paper is organized as follows. Section 2 discusses recent advances with respect to counter measures to 2D facial spoofing attacks. Section 3 introduces our anti-spoofing method for face-based recognition systems and Section 4 shows the experiments performed to validate the proposed solution with respect to the state-of-the-art. Finally, Section 5 concludes the paper and discusses some future research directions.

2. State-of-the-Art

As we mentioned in Section 1, we categorize current anti-spoofing methods into the non-disjoint groups: data-driven characterization, user behavior modeling with respect to the sensor, need of user cooperation, and presence of anti-spoof specific devices. For a more in-depth discussion, Pan et al. [15] and Nixon et al. [13] present good surveys on the biometrics spoofing literature.

Considering the group of data-driven characterization methods, some anti-spoofing techniques for facial recognition systems rely on Fourier analysis. Some researchers explored the high frequencies of Fourier spectra in order to collect features to differentiate between live faces and certain types of spoofs, such as printed images [11, 24].

Other used data-driven approaches include the surface texture of the facial skin from which we can calculate certain measures to characterize optical qualities of the facial skin of live people and compare to non-live ones [16] and optical-flow analysis [1, 8]. Assuming the region of analysis as a 2-D plane, Bao et al. [1] obtained a reference field from the actual flow field data on live and non-live images pointing out their differences. Another solution based on optical-flow analysis was presented by Kollreider et al. [9]. In their work, the authors described two approaches: one using a data-driven characterization that estimates the face motion based on optical flow analysis over selected frames and a second solution exploring a model-based local Gabor decomposition used in conjunction with SVM experts for face part detection.

For the group of approaches counting on the user behavior in front of the camera, some researchers have focused on motion detection such as eye blinking [12, 14] and involuntary movements of parts of the face and head [9, 15]. Kollreider et al. [10] introduced a technique for motion analysis with applications for spoofing detection using the notion of quantized angle features ("quangles") and machine learning classifiers.

One problem with some of the previously mentioned approaches is that they are still impacted by small head tilts which simulate head movement or by short video sequences displaying an authentic user. If we count on the user behavior and also require his/her involvement, we can take advantage of multi-modal information (e.g., voice or gesture) and various challenge-response methods such as asking the user to blink the eyes in a given order, or even smile [5, 13].

If we are allowed to use specific anti-spoof hardware, we can deploy near infrared or thermal images [23]. We can also use 3-D cameras or multiple 2-D cameras to provide additional protection [6].

Although it is clear that important advances have been done regarding spoofing detection such as the aforementioned ones, this research topic is still an open problem. Two important challenges nowadays refer to: (1) the need for designing non-intrusive methods without extra devices and human involvement; and (2) designing detection methods robust to changes in pose and illumination. In this paper, we deal with the first problem presenting a data-driven solution in which we are able to collect several important low-level features directly from the available face data and automatically weight them using partial least squares. We also partially tackle the second challenge by coping with small pose and illumination variations, as we shall discuss in Section 3 for the data sets we consider in this paper.

3. Anti-Spoofing Proposed Solution

A careful observation of the facial spoofing attack samples provides some insights regarding the characteristics that can be explored to design a classification method. In a real access to the system, the person is able to perform slight movements with the head as well as there may exist eye blinking. On the other hand, in an attempt of attack, since a picture is being used, the movements of the head are not independent from the background, the face and the background are in the same plane, there is no eye blinking, and the quality of the printed photo might be a clue by itself.

It is valuable, therefore, to explore both spatial and temporal information to learn differences between live and non-live faces. This suggests the use of a discriminative approach able to locate the most discriminative regions around the face. Our solution employs a holistic representation of the face region through a robust set of low-level feature descriptors, so that differences between classes can be estimated directly in the feature space, which is less prone to variations resulting from uncontrolled acquisition condi-

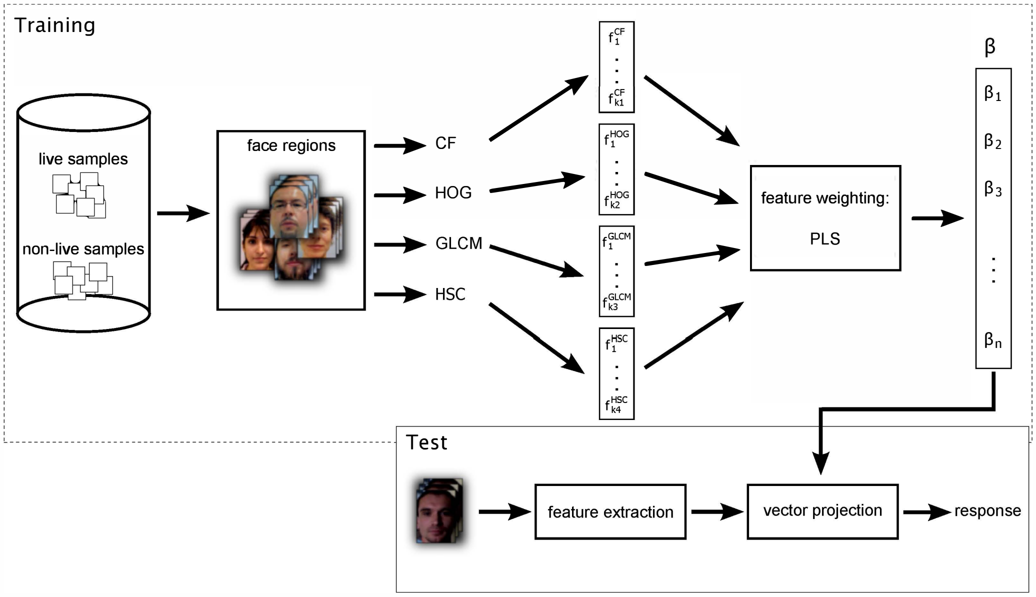

Figure 1: Devised solution to face spoofing detection. Given a set of examples and counter-examples (videos or images) of face spoofing attack, feature vectors (composed by a combination of feature descriptors extracted from the facial region) representing both classes are used to obtain a weighting based on partial least squares, so that novel samples can be classified during the test stage.

tions [21], common in this domain.

Given that a holistic representation is being considered without explicit modeling of the characteristics to be captured (e.g., head movements and eye blinking), it is important to use a robust description of the samples so that models dependent on the application domain can be estimated. Such a description can be obtained with the combination of feature descriptors focusing on different image characteristics, such as shape, color, and texture [22].

To take advantage of the rich information provided by multiple feature descriptors, the anti-spoofing proposed solution integrates feature descriptors based on histogram of oriented gradients (HOG) [4], color frequency (CF) [22], gray level co-occurrence matrix (GLCM) [7], and histograms of shearlet coefficients (HSC) [20] with a weighting scheme based on partial least squares (PLS) [26]. The training and testing procedures are illustrated in Figure 1 and discussed in more details in the next paragraphs.

3.1. Low-Level Descriptors and Feature Extraction

Since our method focuses on the facial regions, a face detector based on [22] is first applied to the samples, which are cropped and resized to a common size so that the feature extraction can be performed. The feature descriptors considered in this paper include information related to shape (histogram of oriented gradients and histograms of shearlet coefficients), color (color frequency), and texture (gray level co-occurrence matrix). A summary of these descriptors is presented as follows.

HOG captures edge or gradient structures that are characteristic of local shape. According to Dallal and Triggs [4], a consequence is a controllable degree of invariance to local geometric transformations, in which it presents invariance to translations and rotations smaller than the local spatial or orientation bin size. The HOG employed in this work also considers the color frequency (CF) descriptor [22], responsible for capturing color information.

HSC analyzes edges at multiple scales and orientations based on the multi-scale analysis provided by shearlet transforms. Histograms are employed to estimate the edge response distribution at each decomposition level provided by the shearlet transform. At the end, the histograms resulting from each decomposition level are concatenated and normalized to be used as a feature descriptor.

To capture texture properties, we extract features from co-occurrence matrices that represent the joint probability distribution of gray-level pairs of neighboring pixels in a block. After calculating the co-occurrence matrices of a given image for four orientations, 12 descriptors are extracted: angular second-moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, and directionality [7] to summarize such matrix.

The input samples to the system can be either videos or images depending on the type of attack under consideration. In the former case, aiming at exploiting both temporal and spatial information, a video sample containing n frames is divided into m parts, such that the feature extraction is performed for every k-th frame (to avoid extremely high dimensional feature spaces), where $k = \lfloor n/m \rfloor$. The re-

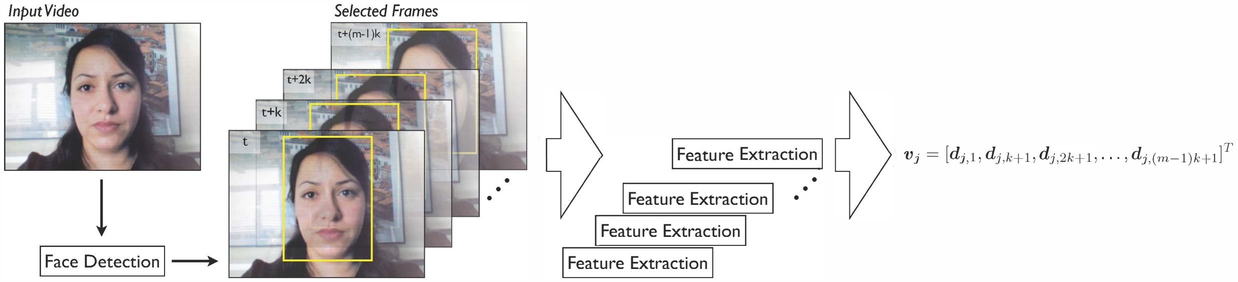
Input Video
Selected Frames
t+(m-1)k
t+2k
t+k
t
Face Detection
Feature Extraction
Feature Extraction
Feature Extraction
Feature Extraction
$\boldsymbol{v}_j = [\boldsymbol{d}_{j,1}, \boldsymbol{d}_{j,k+1}, \boldsymbol{d}_{j,2k+1}, \ldots, \boldsymbol{d}_{j,(m-1)k+1}]^T$

Figure 2: Feature extraction process for an input video. A video containing n frames is divided into m parts, such that the feature extraction is performed for every k-th frame. The resulting feature vector, $\boldsymbol{v}_j$, is composed by concatenating descriptors extracted from each frame.

sulting descriptors are concatenated to compose the feature vector used to describe the video sample, as illustrated in Figure 2. In the latter case, when image samples are considered, the same procedure is performed for the special case where $n = m = 1$.

The feature extraction for the t-th frame of the j-th sample (after face detection, cropping, and resizing) is performed as follows. The frame is split into overlapping blocks with different sizes and strides (being able to capture more visual information), then the descriptors extracted from each block are concatenated creating a feature vector $\boldsymbol{d}_{j,t}$. Finally, when a video sample has descriptors extracted from all its selected frames, a high-dimensional feature vector $\boldsymbol{v}_j = [\boldsymbol{d}_{j,1}, \boldsymbol{d}_{j,k+1}, \boldsymbol{d}_{j,2k+1}, \ldots, \boldsymbol{d}_{j,(m-1)k+1}]^{\mathrm{T}}$ is composed to describe the j-th sample.

3.2. Partial Least Squares Regression

The use of a robust set of feature descriptors renders many classical machine learning methods intractable due to (1) the extremely large resulting feature space, which becomes even more evident when the temporal information is considered (descriptors are extracted from multiple frames); (2) the reduced number of training samples compared to the number of descriptors, and (3) the increase of the multicollinearity among the training samples. However, the statistical method called Partial Least Squares [26] is not affected by such problems [19].

Partial least squares is a method for modeling relations between sets of observed variables in a latent space. It constructs new predictors as linear combinations of the original variables summarized in a matrix $\boldsymbol{X}$ of descriptor variables (matrix with feature vectors) and a vector $\boldsymbol{y}$ of responses (training class labels). PLS decomposes the input variables as

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}$$
$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{q}^T + \boldsymbol{f}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are $n \times p$ matrices containing p extracted latent vectors. The $(m \times p)$ matrix $\boldsymbol{P}$ and the $(1 \times p)$ vector $\boldsymbol{q}$ represent the loadings. The $n \times m$ matrix $\boldsymbol{E}$ and the $n \times 1$ vector $\boldsymbol{f}$ are the residuals. The PLS method, using the nonlinear iterative partial least squares (NIPALS) algorithm [26], constructs a matrix of weights $\boldsymbol{W}$ indicating the importance of each descriptor. Using these weights, the regression coefficients $\boldsymbol{\beta}_{m \times 1}$ can be estimated by

$$\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{y}. \tag{1}$$

The regression response for feature vector $\boldsymbol{v}_j$ is obtained by

$$y_j = \overline{y} + \boldsymbol{\beta}^T\boldsymbol{v}_j \tag{2}$$

where $\overline{y}$ is the sample mean of $\boldsymbol{y}$.

When PLS is employed to obtain the latent feature space, higher weights are attributed to feature descriptors located in regions containing discriminatory characteristics between the two classes. Figure 3 depicts an example of the weight distribution on the face region obtained during the training stage by PLS considering HOG descriptors. The importance of regions around the eyes and nose to discriminate between live and non-live samples is clear.

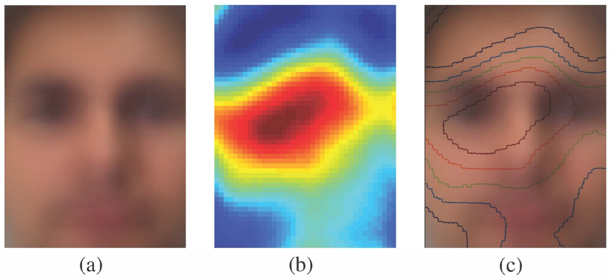
(a) (b) (c)

Figure 3: Spatial distribution of the feature descriptors weighting obtained by PLS. (a) average face considering training samples of FSA dataset; (b) feature weighting map, where red indicates high discriminative power and blue low; (c) average face overlaid with contour curves extracted from the feature weighting map.

3.3. Training and Testing Procedures

The procedure to estimate a PLS regression model to face spoofing detection is illustrated on top of the diagram depicted in Figure 1. Given a set of live, $S_l =$

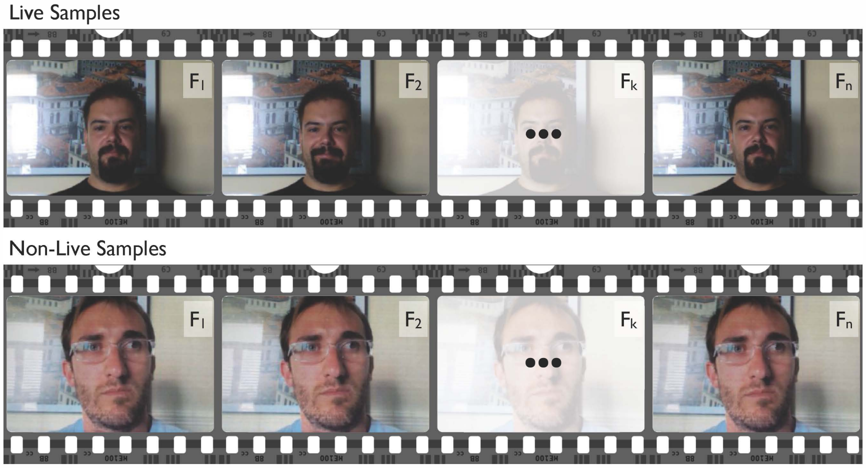

Figure 4: Samples of the FSA dataset.

$\{s_{l1}, s_{l2}, \ldots, s_{lo}\}$, and non-live, $S_n = \{s_{n1}, s_{n2}, \ldots, s_{np}\}$, training samples (images or videos according to the dataset), the training process is described as follows.

Once the faces in sets S_l and S_n are detected, cropped and rescaled to a common size (so that the feature extraction can be performed with samples nearly registered), descriptors are extracted from a selected number of frames using CF, HOG, HSC, and GLCM, and then concatenated to compose a feature vector, as illustrated in Figure 2. This process results in two matrices with feature vectors on the columns $V_l = [\boldsymbol{v}_{l1}, \boldsymbol{v}_{l2}, \ldots, \boldsymbol{v}_{lo}]$ and $V_n = [\boldsymbol{v}_{n1}, \boldsymbol{v}_{n2}, \ldots, \boldsymbol{v}_{np}]$, representing the live and non-live classes, respectively.

From matrix $X = [V_l, V_n]^{\mathrm{T}}$ and the response vector $\boldsymbol{y}$ with its first o elements equal to $+1$ and its last p elements equal to -1, indicating the sample class labels, the PLS regression model can be learned. The resulting $\boldsymbol{\beta}$ regression coefficients, estimated according to Equation 2, are stored to be used during the test to evaluate the class of an unseen sample.

Finally, the test procedure evaluates if a novel sample belongs either to the live or non-live class. When a sample video is presented to the system, the face is detected and the frames are cropped and rescaled. Then, the vector $\boldsymbol{v}_j$, resulting from the feature extraction, is projected onto $\boldsymbol{\beta}$, responses close to +1 indicate live samples and responses close to -1 indicate non-live samples.

4. Experiments and Validation

This section evaluates several aspects of the proposed approach on datasets based on video and image attack attempts. First, we perform experiments considering the FSA video dataset to compare the achieved results to other teams participating of the *2011 IJCB Competition on Counter Measures to 2D Facial Spoofing Attacks*, after assessing the influence of the number of selected frames and the combination of feature descriptors. Then, we compare the proposed solution to previously published results using the image-based NUAA dataset [24].

4.1. Evaluation on the FSA Dataset

The FSA dataset, developed for the *2011 IJCB Competition on Counter Measures to 2D Facial Spoofing Attacks*, consists of 200 real-access attempt (live) and 200 printed-photo attack attempt (non-live) videos, with different length and lighting conditions. The attack attempts can be performed with fixed or hand-held printed photos. Figure 4 shows samples of this dataset.

The FSA dataset is divided into three sets: training, development and test. The first two sets have 60 videos for each class and the last has 80 samples per class. In our experiments, the development set is used to estimate a set of suitable parameters that will be used to perform spoofing detection in the test samples.

Experimental Setup. Before performing feature extraction, the faces are detected using the detector proposed in [22] and rescaled to 110×140 pixels so that all images present the same size. Thereafter, feature extraction is performed for the selected frames of each video using the following parameters. For HOG and CF feature descriptors, eight orientations are considered with block sizes of 32×32 and 16×16 with strides of eight and four pixels, respectively. For HSC, with the same block sizes and strides used by HOG, two decomposition levels and eight orientations are considered. For GLCM we used blocks of 32×32 and 16×16 with strides of eight and pixels each. The final feature vector length extracted for each frame is $109,460$.

Number of Selected Frames. This experiment evaluates how the number of frames selected to represent a video sample affects the equal error rate (EER). On one hand, the more frames are used the more information is captured over

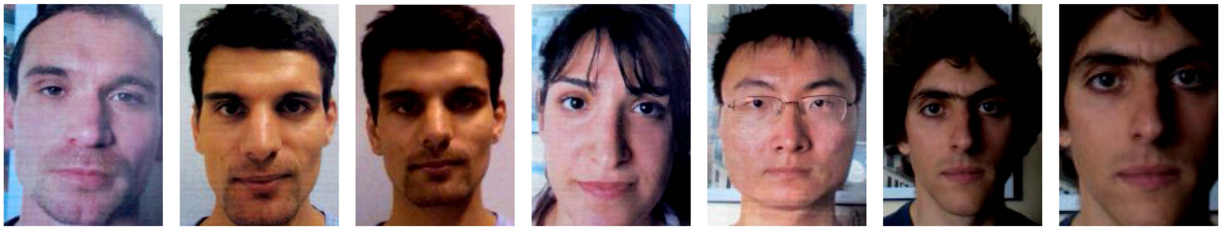

Figure 5: Misalignment of detected faces in the FSA dataset.

time; on the other hand, the dimensionality of the feature space becomes prohibitively large.

For this experiment, we consider only the HOG feature descriptor and evaluate the average and standard deviation of the EER as a function of the number of selected frames. The number of frames considered varied between 1 and 13. For each number, the detection was executed 10 times, each one with a different initial frame. Results showed that the EER becomes stable when the number of selected frames is equal or greater than 10 (standard deviation for the EER approaches zero). Therefore, in the remaining experiments with the FSA dataset, we use the number of selected frames per video equal to 10.

Combination of Feature Descriptors. This experiment assesses improvements achieved when feature descriptors are combined. Table 1 shows the equal error rate obtained in the development set when feature descriptors are considered individually and combined. These results raise three main points, discussed as follows.

Name	# descriptors	EER (%)
HOG	326,880	11.67
Intensity	154,000	8.33
CF	27,240	6.67
GLCM	159,360	6.67
HSC	581,120	4.33
combination	1,094,600	1.67

Table 1: Equal error rate achieved by single feature descriptors and their combination considering block-based feature descriptors (HOG+CF+GLCM+HSC).

First, the use of the pixel intensity does not provide enough information to achieve low equal error rates. This is mainly due to incorrect pixel registration, once an automatic detector is employed to extract the faces and no further alignment is performed, as Figure 5 shows. These misalignments emphasize the need for block-based feature descriptors since they are robust to small local variations.

Second, as pointed out earlier, the quality of the printed photos can help discriminating between live and non-live samples, as Figure 4 depicts, in which the non-live sample presents subtle horizontal color lines. This is supported by the fact that descriptors capturing color (CF) and texture (GLCM) information performed better than shape-based HOG descriptor. However, even though the HSC descriptor is also based on shape, its multi-scale approach captures important discriminative information.

Finally, the combination of block-based feature descriptors (HOG+CF+GLCM+HSC) provided the best results in the development set, reducing the EER in about 2.5 times compared to the results achieved by HSC, the best individual feature descriptor. Therefore, analyses of color, texture, and multi-scale provide complementary information improving face spoofing detection.

Classifier Evaluation. This experiment evaluates the use of PLS and SVM to classify samples from the development set considering the feature combination (feature vectors with $1,094,600$ variables). The data was first standardized to zero mean and standard deviation equal to one. The SVM implementation used was libSVM [3] with type C and a linear kernel[1]. For the PLS method, the EER is equal to 1.67% (Table 1) and with SVM the EER is 10%.

	Development	Test	
Research team	EER (%)	FAR (%)	FRR (%)
IDIAP	0.00	0.00	0.00
UOULU	0.00	0.00	0.00
AMILAB	0.00	0.00	1.25
CASIA	1.67	0.00	0.00
SIANI	1.67	0.00	21.25
our team	1.67	1.25	0.00

Table 2: Results achieved by teams participating of the 2011 IJCB Competition on Counter Measures to 2D Facial Spoofing Attacks in the development and test sets. Other participating teams: Idiap Research Institute (IDIAP), Switzerland; Machine Vision Group (UOULU), University of Oulu, Finland; Ambient Intelligence Laboratory (AMILAB), Italy; Center for Biometrics and Security Research, Institute of Automation, Chinese Academy of Sciences (CASIA), China; and Universidad de Las Palmas de Gran Canaria, (SIANI), Spain.

Results and Comparisons. Table 2 shows the official results of the *2011 IJCB Competition on Counter Measures to*

[1] The experiments were performed on a 64 bits OS with 12GB of memory. The memory was not enough to run this experiment with non-linear SVM kernels.

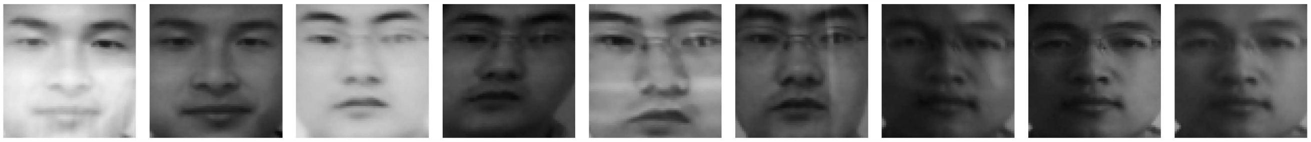

Figure 6: Samples of the NUAA dataset after rescaling and normalization.

2D Facial Spoofing Attacks [2]. The experiment was conducted as follows. First, the threshold for the EER on the development set was obtained, then it was applied for discriminating samples in the test set, resulting in values of false accept rate (FAR) and false reject rate (FRR) shown in the table.

According to Table 2, most of the teams achieved very small error rates in both sets. In the development set, at most two video samples were incorrectly classified by all methods out of 120. In the test set, our team incorrectly classified only two out of 160 videos.

4.2. Evaluation on the NUAA Dataset

The NUAA dataset, proposed by Tan et al. [24], comprises images extracted from videos of 15 subjects captured in three sections and contains attempts of attack based on hand-held printed photos. This dataset is divided into training and test sets. The former has $1,743$ live images and $1,748$ non-live, and the latter consists of $3,362$ live and $5,761$ non-live samples.

The data provided in this dataset consists of images with faces cropped using the Viola-Jones detector [25] and gray-scale images normalized to 64×64 pixels aligned by the nose and eyes. Tan et al. [24] used the normalized images in their experiments. To perform a direct comparison to their results, we also use the same set of images.

Experimental Setup. Since the images used in the experiments are gray-scale, the color frequency feature was not employed. The remaining parameters are the same used with the FSA dataset. The final length of the feature vector to describe each sample is $22,952$.

In the following experiments, equal error rate (EER) and the area under the ROC curve (AUC) are used to show the results achieved in the NUAA dataset. In the AUC, the larger the value achieved, the better the results.

Combination of Feature Descriptors. According to Table 3, the feature combination improved spoofing detection. This result is in accordance with the ones obtained with the FSA dataset. In addition, it is worth pointing out the poor quality of results achieved by the intensity alone, which is a consequence of the strong illumination changes present on the NUAA dataset (Figure 6). This problem is reduced when other feature descriptors are employed.

Results and Comparisons. Figure 7 shows the ROC curve achieved by the proposed method with feature combination (HOG+GLCM+HSC). For the same data, the best result reported by Tan et al. [24] achieved AUC of 0.95.

Name	# descriptors	EER (%)	AUC
Intensity	4,096	52.20	0.425
HOG	6,984	16.80	0.908
HSC	12,416	12.40	0.944
GLCM	3,552	9.60	0.960
combination	22,952	8.20	0.966

Table 3: Equal error rate and area under the ROC curve achieved by single feature descriptors and their combination considering block-based feature descriptors (HOG+GLCM+HSC).

5. Conclusions

This paper introduced an anti-spoofing solution based on a set of low-level feature descriptors exploring both spatial and temporal information using Partial Least Squares regression to provide a feature weighting to distinguish between 'live' and 'spoof' images or videos. The devised facial anti-spoofing solution worked well in both evaluated datasets (for video and image data) without changes of parameters other than the adaptation of the feature descriptors to gray-scale images that compose the NUAA dataset.

Since it is difficult to know beforehand which feature descriptors are suitable to perform spoofing detection for a given dataset (e.g., for the FSA dataset, the HSC performs better, however, the GLCM provides better results for the NUAA dataset), the use of PLS regression allowed the combination of multiple feature descriptors even though the resulting feature space is extremely high dimensional, without the need of choosing a subset of features in advance. In addition, the results showed that the combination performed by PLS provided the best results for both datasets when compared to the use of individual features.

As seen in Figure 5 and in the high EER achieved with pixel intensity shown in Table 1, the misalignment of the faces automatically detected is responsible for loosing some accuracy in the spoofing detection. Therefore, further detection improvements might be obtained with the addition of a module to register the faces according to the eyes and nose.

It is also important to note that the results achieved by our method in the NUAA dataset are based only on features considering shape and texture since we used the normalized gray-scale images. Additional visual information

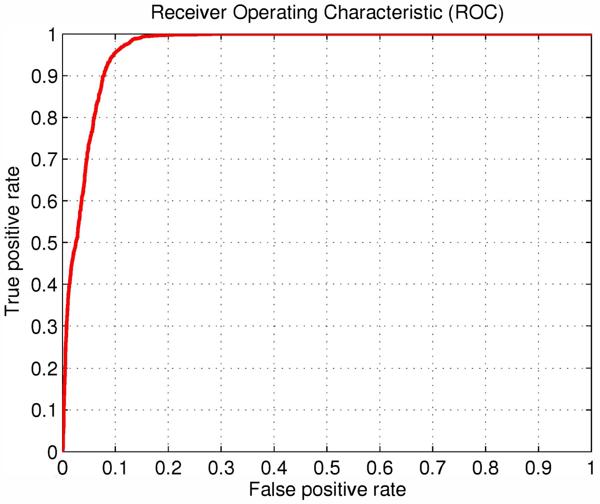

Figure 7: ROC obtained with feature combination for the NUAA dataset. AUC = 0.966.

could have been extracted if the higher resolution detected color faces were used. However, the annotations provided by the creators for the nose and eye locations are not accurate for a large number of samples, which leads to severe misalignments of the faces.

Acknowledgments

The authors are thankful to FAPESP, CNPq, Microsoft and CAPES for the financial support. This research was partially supported by FAPESP grants 2010/10618-3 and 2010/05647-4.

References

[1] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In *IEEE IASP*, pages 233–236, 2009.

[2] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Li, W. R. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. Castrillon-Santana, and J. Maatta. Competition on Counter Measures to 2D Facial Spoofing Attacks. In *IJCB*, 2011.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE CVPR*, pages 886–893, 2005.

[5] European Commission BioSecure Project. Biometrics for secure authentication. Technical Report, European Commission BioSecure, 2007. http://biosecure.it-sudparis.eu/AB/.

[6] T. Fladsrud. Face recognition in a border control environment: Non-zero effort attacks "effect on false acceptance rate". Msc. thesis, Gjovik Univ. College, Norway, 2005.

[7] R. Haralick, K. Shanmugam, and I. Dinstein. Texture Features for Image Classification. *IEEE TSMC*, 3(6), 1973.

[8] K. Kollreider, H. Fronthaler, and J. Bigun. Evaluating liveness by face images and the structure tensor. In *IEEE AutoID*, pages 75–80, 2005.

[9] K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. *Elsevier IVC*, pages 233–244, February 2009.

[10] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun. Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE TIFS*, 2(3):548–558, 2007.

[11] J. Li, Y. Wang, T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra. In *Biometric Technology for Human Identification*, pages 296–303, 2004.

[12] J.-W. Li. Eye blink detection based on multiple gabor response waves. In *IEEE ICMLC*, pages 2852–2856, 2008.

[13] K. Nixon and V. A. R. Rowe. *Handbook of Biometrics*, chapter Spoof Detection Schemes. Springer, 2008.

[14] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *IEEE ICCV*, pages 1–8, 2007.

[15] G. Pan, Z. Wu, and L. Sun. *Recent Advances in Face Recognition*, chapter Liveness detection for face recognition, pages 235–252. InTech, 2008.

[16] G. Parziale, J. Dittmann, and M. Tistarelli. Analysis and evaluation of alternatives and advanced solutions for system elements. *BioSecure*, 2005.

[17] B. Peixoto, C. Michelassi, and A. Rocha. Face liveness detection under bad illumination conditions. In *IEEE ICIP*, 2011.

[18] F. L. Podio. Biometrics technologies for highly secure personal authentication. ITL Bulletin, Information Technology Laboratory, NIST, May 2001.

[19] R. Rosipal and N. Krämer. *Subspace, Latent Structure and Feature Selection*, chapter Overview and Recent Advances in Partial Least Squares, pages 34–51. Springer, 2006.

[20] W. R. Schwartz, R. D. da Silva, and H. Pedrini. A Novel Feature Descriptor Based on the Shearlet Transform. In *IEEE ICIP*, 2011.

[21] W. R. Schwartz, H. Guo, and L. S. Davis. A Robust and Scalable Approach to Face Identification. In *ECCV*, volume 6316 of *Lecture Notes in Computer Science*, pages 476–489, 2010.

[22] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In *IEEE ICCV*, pages 24–31, 2009.

[23] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel. Face recognition with visible and thermal infrared imagery. *Elsevier CVIU*, pages 72–114, July 2003.

[24] 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*, pages 504–517, 2010.

[25] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004.

[26] H. Wold. Partial least squares. In S. Kotz and N. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 6, pages 581–591. Wiley, New York, 1985.