IDIAP RESEARCH REPORT

COMPETITION ON COUNTER MEASURES TO 2-D FACIAL SPOOFING ATTACKS

Murali Mohan Chakka André Anjos
Sébastien Marcel

Idiap-RR-29-2011

AUGUST 2011

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Competition on Counter Measures to 2-D Facial Spoofing Attacks

Murali Mohan Chakka[1], André Anjos[1], Sébastien Marcel[1], Roberto Tronci[2], Daniele Muntoni[2], Gianluca Fadda[2], Maurizio Pili[2], Nicola Sirena[2], Gabriele Murgia[2], Marco Ristori[2], Fabio Roli[2], Junjie Yan[3], Dong Yi[3], Zhen Lei[3], Zhiwei Zhang[3], Stan Z.Li[3], William Robson Schwartz[4], Anderson Rocha[4], Helio Pedrini[4], Javier Lorenzo-Navarro[5], Modesto Castrillón-Santana[5], Jukka Määttä[6], Abdenour Hadid[6], Matti Pietikäinen[6]
Idiap Research Institute[1], Ambient Intelligence Laboratory[2], Chinese Academy of Sciences[3], University of Campinas[4], Universidad de Las Palmas de Gran Canaria[5], University of Oulu[6]
{murali.chakka,andre.anjos,sebastien.marcel}@idiap.ch[1],
{roberto.tronci,muntoni,fadda,maurizio.pili,sirena,gabriele.murgia,ristori}@sardegnaricerche.it[2],
roli@diee.unica.it[2], {jjyan,dyi,zle,zwzhang}@cbsr.ia.ac.cn[3], stan.zq.li@gmail.com[3],
{schwartz,anderson.rocha,helio}@ic.unicamp.br[4], {jlorenzo,mcastrillon}@iusiani.ulpgc.es[5],
{jukmaatt,hadid,mkp}@ee.oulu.fi[6]

Abstract

Spoofing identities using photographs is one of the most common techniques to attack 2-D face recognition systems. There seems to exist no comparative studies of different techniques using the same protocols and data. The motivation behind this competition is to compare the performance of different state-of-the-art algorithms on the same database using a unique evaluation method. Six different teams from universities around the world have participated in the contest. Use of one or multiple techniques from motion, texture analysis and liveness detection appears to be the common trend in this competition. Most of the algorithms are able to clearly separate spoof attempts from real accesses. The results suggest the investigation of more complex attacks.

1. Introduction

Face recognition has been an active research topic in the last two decades and its techniques are currently deployed in access control systems. Facial recognition has the advantage of non-intrusiveness over the other biometric identification techniques such as irises and finger prints. However, spoofing attacks is a major threat causing problems to face recognition to be used as a biometrics for high-security applications.

The use of facial photographs of a valid user to spoof face recognition is the most common attack method, as the photographs of the users are widely available through websites like social networks. Even videos of the users can be easily captured from distant cameras without prior consent. To make face recognition as a successful biometric identification technology, there exist the necessity of answering the spoofing attack problem .

Based on the clues used for attack detection, anti-spoofing techniques for 2-D face recognition can be roughly classified as *motion, texture* and *liveness*. *Motion analysis* techniques use the fact that, planar objects move significantly different from real human faces which are 3-D objects. Kollreider *et al.* [13] evaluate the trajectories of selected part of the face from the short sequence of images using a simplified optical flow analysis followed by a heuristic classifier. The same authors introduce a method [14] to fuse these scores with liveness properties such as eye-blinks or mouth movements. Bao *et al.* [2] propose a method to detect attacks produced with planar media using optical flow based motion estimation.

Texture analysis techniques take the advantage of detectable texture patterns such as printing failures, and overall image blur to detect attacks. Li *et al.* [15] detect print-attacks by exploiting differences in the 2-D Fourier spectra of hard-copies of faces and real-accesses. The method works well for down-sampled photo attacks, but is likely to fail for higher-quality samples. Bai *et al.* [1] analyze the micro-textures using a linear SVM classifier to detect spoof attacks.

Liveness detection tries to classify attacks based on the signs of life such as eye-blinks and mouth-

movements. Pan *et al*. [18, 19] bring a real-time liveness detection specifically against photo-spoofing using eye-blinks.

In spite of several advances in anti-spoofing for face recognition, there seems to exist no comparative studies of different techniques on a publicly available database. Therefore, the motivation behind this competition is to compare the performance of different state-of-the-art algorithms on the same database using a unique evaluation method. Six different teams from universities world-wide had participated in the contest, they are *Ambient Intelligence Laboratory* (AMILAB), Italy; *Center for Biometrics and Security Research, Institute of Automation, Chinese Academy of Sciences* (CASIA), China; *Idiap Research Institute* (IDIAP), Switzerland; *Universidad de Las Palmas de Gran Canaria*, (SIANI), Spain; *Institute of Computing* (UNICAMP), Brazil; and *Machine Vision Group (UOULU), University of Oulu*, Finland.

In Section 2, we briefly review the database and evaluation protocols used for the competition. Section 3 presents the algorithms of all six participants. We discuss the consolidated results of all algorithms in Section 4 and finally conclude the paper in Section 5.

2. Database & Protocols

For the competition, we used the publicly available PRINT-ATTACK biometric (face) database[1]. The database consists of 200 videos of real accesses and 200 videos of attack attempts of 50 different identities. Real access video sequences are captured with 320 by 240 pixel (QVGA) resolution, at 25 frames-per-second and for 15 Seconds duration each. These real access videos are recorded under two different background and illumination conditions. Attack attempts are captured with the same resolution and frame rate, for up to nine seconds duration under the same background and illumination conditions. Hard copies of the digital photographs printed on plain A4 paper using color laser printer are used for recording attacks. For attacks, two different support mechanisms are installed in-front of the input camera of the acquisition system. The supports used are, *hand-based* in which an attacker holds the client's print in his hands and *fixed-support* in which the print is attached to the wall.

In the competition, contestants were given access to training and development data sets, each set containing 60 real accesses and 60 attack attempts. The sample identities for these data sets were drawn randomly without repetition. All the teams were given a couple of months to train and develop their classification system. For training and development, participants were free to use the entire video of each client, 375 frames for real accesses and 230 for attack attempts. Then we released the test data set containing 80 real access and 80 attack videos. All videos in the test set contained 230 frames. Files of the test data set are anonymized to conceal the type of the video (real/attack, hand-based/fixed-support) for true evaluation. Every participant's algorithm is supposed to yield a score after processing 230 frames of each video in the test data set. We had asked all the teams to provide two files containing such scores, one for the development set and the other for the test set. We were not interested in speed, latency and complexity of the contestant's method.

In order to compute the performance measure of the spoofing detection systems, we first computed a threshold at Equal Error Rate (EER) on the development set scores. Then, on the test set scores using the same threshold, we computed Half Total Error Rate (HTER), which combines the False Rejection Ratio (FRR) and the False Acceptance Ratio (FAR) with 0.5 weight. Spoof detection accuracy is good if FAR/FRR/HTER value is close to zero percent.

3. Methods

In this section, we consolidate the algorithms proposed by the participants of the competition.

3.1. AMILAB

We faced the problem of detecting 2-D face spoofing attacks performed by placing a printed photo of a real user in front of the camera. Unfortunately, it is not possible to relay just on the face movement as a clue of vitality because the attacker can easily simulate such a case and also because real users often show a "lower vitality" during the authentication session. Therefore, our approach consists of performing both video and still image analysis in order to employ complementary information about vitality and consequently to obtain a more robust classification.

From our experience, an image analysis performed over videos shows clear peculiar visual characteristics for captured printed photos and real scenes. To detect the differences we explored several different types of visual features (e.g. color features, edges, textures etc.), and we used a set of support vector machines (SVMs) to compute a frame level confidence score of being a real session or not. To obtain an high separation between score distributions we combined these similarity scores by means of the Dynamic Score Combination methodology [25]. This type of analysis can be performed also frame by frame.

Even if the peculiarities described above can be detected, the vitality detection must be used to assess

[1] `http://www.idiap.ch/dataset/printattack`

a certain degree of scene reality. We computed two more vitality scores on videos: the first one is based on the average movement caught by a common motion detection technique [16], the second one depends on the number of eye blinks [4]. Blinks are a proof of vitality, but nothing can be inferred in case of their absence. Conversely low degrees of movement imply an attack but movement is not distinctive for hand attacks and real sessions.

Combination was performed at score level as a weighted sum: photo detection provided excellent results, therefore it was given a higher weight in combination; movement measures provided some contribution only for fixed photo attacks therefore the score is used only in case of very little movement. By taking into account the above considerations, we used the following score combination scheme:

$$s^* = \begin{cases} \alpha \cdot s_{image\ analysis} + (1-\alpha) \cdot s_{blinks}, \\ \qquad\qquad \text{if } s_{motion} \text{ is high} \\ \alpha_1 \cdot s_{image\ analysis} + \alpha_2 \cdot s_{blinks} + \alpha_3 \cdot s_{motion}, \\ \qquad\qquad \text{if } s_{motion} \text{ is low} \end{cases}$$

3.2. CASIA

Our method is based on the following intuitive three observations: (1) Real access videos tend to have non-rigid motions, especially in the eye and mouth regions, while printed photos only have rigid transformations like translation, scaling, and rotation; (2) Real access videos tend to have less noise than those spoofing videos; (3) Real access videos only have local motions in the face region while spoofing videos usually have global motions spread-out the whole support. We analyze these three clues and construct three classifiers based on them respectively.

Classifier 1 - Non-Rigid Motion: The non-rigid motions in real faces are detected by a batch image alignment technique proposed in [21], called "RASL". Since the spoofing videos are generated by the fixed or hand-shaking printed photos, the motion of these videos can be well modeled by affine transformation. RASL algorithm has the ability to align a series of affine transformed images, therefore the frames of spoofing videos can be well aligned. For real faces, due to the non-rigid motions, there still be having large variations in the aligned frames. For this reason, the differences of the aligned frames are used to construct the first classifier.

Classifier 2 - Noise: The spoofing videos usually are noisier than those in real faces. Following the methods in [8] and [7], noise variance can be estimated by a robust median estimator as follows

$$\hat{\sigma} = \frac{Median\{|y(i,j)|\}}{0.6745} \quad y(i,j) \in HH1 \qquad (1)$$

where HH1 means the first order wavelet decomposition of a image. In our experiments, we have tested several wavelet basis functions and find that the Haar wavelet is efficient to evaluate the noise in images. The noise differences between the real and spoofing videos are used to construct the second classifier.

Classifier 3 - Face-Background Dependency: Generally, the motion of a real face is independent of the background. On the contrary, the background around the face is usually moving together with the face in printed photo attacks. This is another significant feature to differentiate the real and spoofing videos.

Specifically, we use the GMM background modeling method [12] to detect the background and Viola-Jones face detector [26] to detect the face area. The ratio of motion in the face region and background is used to evaluate the face-background dependency, which is used to construct the third classifier.

Classifiers Fusion: In our system, the three classifiers are learned by logistic regression respectively. Classifier 1 and classifier 3 are fused to predict scores of videos with complex background, and classifier 1 and classifier 2 are used to predict scores of videos with uniform background. Since the background condition can be easily classified by edge detection, all the procedures are automatic.

3.3. UNICAMP

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 (even slight ones) between the two classes. This suggests the use of an approach able to locate discriminative regions within 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 conditions [23], 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 blink-

ing), it is important to use a robust description of the samples so that models dependent on the application domain can be expendable. Such a description can be obtained with the combination of feature channels focusing on different image characteristics, such as shape, color, and texture [24].

The anti-spoofing proposed solution integrates feature descriptors based on histogram of oriented gradients (HOG) [6], gray level co-occurrence matrix (GLCM) [11], and histograms of shearlet coefficients (HSC) [22] with a weighting scheme based on partial least squares (PLS) regression [27].

To exploit both temporal and spatial information, a sample video containing n frames is divided into m parts, such that the feature extraction is performed for every k-th frame, where $k = \lfloor n/m \rfloor$. 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 to create 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.

To estimate a PLS regression model, we use a set of real access and attempt of attack training samples $S_r = \{s_{r1}, s_{r2}, \ldots, s_{ro}\}$ and $S_a = \{s_{a1}, s_{a2}, \ldots, s_{ap}\}$, respectively. Once the faces are detected, cropped and rescaled to a common size, descriptors are extracted from a selected number of frames and then concatenated to compose a feature vector. This process results in two matrices, $V_r = [\boldsymbol{v}_{r1}, \boldsymbol{v}_{r2}, \ldots, \boldsymbol{v}_{ro}]$ and $V_a = [\boldsymbol{v}_{a1}, \boldsymbol{v}_{a2}, \ldots, \boldsymbol{v}_{ap}]$ representing the real access and attempt of attack classes, respectively, with feature vectors on their columns.

With the availability of a matrix $X = [V_r, V_a]^{\mathrm{T}}$ and the response vector $\boldsymbol{y}$ with its first o elements equals to $+1$ and its last p elements equals to -1, indicating the sample class labels, the PLS regression model can be learned. The estimated regression coefficients are stored in a vector $\boldsymbol{\beta}$ to be used during the test.

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 $\boldsymbol{\beta}$. The response indicates whether this sample is a real access or an attack attempt.

3.4. IDIAP

Photograph attacks, when executed with hard-copies, may suffer from print artifacts or failures that can be used as counter-measure to spoofing. Our proposed scheme processes each video by accumulating a single Local Binary Pattern (LBP) code histogram [17] with data from every single image in the stream. The histogram is matched against a pre-calculated model, using the χ^2 method as proposed on the same reference to generate a final score.

The input video is first converted into gray-scale before passing through the LBP operator configured to use a radius $R = 1$ and the 8 surrounding pixels. The 2-D outputs of the LBP operator are accumulated in a global histogram with 256 bins for all input images in the sequence. After the video input has finished, the global histogram is compared to a reference histogram for attacks, generated using all images from all videos of attacks available at the training set. The output of the χ^2 statistics is ready for classification without any further treatment.

3.5. SIANI

To solve the task, we have made the assumption that the evolution of the face appearance and its location in the image are important cues to distinguish between real and attack videos. Thus, this approach is based on the detection data collected by the ENCARA2 face detector [3]. The face detector features each sequence as a whole to include the temporal information.

The detection data provided by ENCARA2 whenever a face is located are: the face container, and, if available, the eyes, nose, and mouth locations.

To compute the sequence features, we have analyzed the information given by the facial element locations, and a simple difference image with the previous frame. This analysis provides information of interest based on the face and facial elements motion to detect an attack.

For that reason we have computed basic statistics on the facial element locations and specific areas of the difference image. In this sense, the mean position and variance are computed for each element location.

Also, the difference image is analyzed in some specific areas. These specific areas are defined according to the face container, which divides the image into two areas: the face and the non face areas. Additionally, if both eyes are located, the distance between them is calculated and used to estimate the areas of interest around both eyes and the mouth. In summary, up to five areas of interest are featured: face, non face, eyes and mouth. A measure based on the difference image is computed for all those areas and normalized attending to their respective size.

The classification of the videos between real and attacks is done with the Bayesian Network approach [20] included in the Weka open source software [10]. Among the different algorithms available we have selected the Chow and Liu algorithm [5], that reported the best re-

sults for the *devel* set. The results achieved for both sets allow the possibility of selecting a threshold to have 0% FAR for both *test* and *devel* sets.

3.6. UOULU

Our spoofing detection approach was inspired by image quality assessment and characterization of printing artifacts [9]. It is assumed that the face prints contain printing quality defects which can be recognized using texture features. Based on this, the system can detect whether there is a live person in front of the camera or a face print.

The proposed system considers only single video frames for liveness detection and performs texture analysis on a window containing the face area and its surrounding regions. First, a Sobel horizontal edge-emphasizing filter is applied to highlight the image defects and to produce a gradient image from which a single local binary pattern (LBP) feature histogram is computed [17]. The LBP representation ($LBP^{u2}_{8,2}$) computed over the preprocessed face image and its surrounding regions encodes the occurrences of local texture primitives and describes well the differences between a real face and a printed face image. The 59-bin histogram is fed to an SVM classifier which determines whether or not the window contains a real face.

4. Discussion

Table 1 summarizes the performance of the algorithms of all the participants. We have observed that all teams are using one or multiple clues obtained clearly from three types of techniques – motion analysis, texture analysis and liveness detection. Three teams had achieved HTER of zero percent on test set. Two teams, IDIAP and UOULU have obtained zero percent EER on development set and zero percent HTER on test set based on texture analysis method. This leads to the conclusion that, the attack videos in this database mainly consist of detectable texture patterns. Incidentally both teams use LBP as the base technique to compute scores for classification. The CASIA team has presented a method with the combination of motion and texture analysis techniques, and the method also allows switching between detection schemes based on the scene context. Even though this method has reported perfect detection only on test set, the scheme appears to be robust, but at the expense of complexity. The AMILAB and the UNICAMP teams used all three experts in deriving the detection scheme. The AMILAB team has reported near-perfect result with only one false rejection on the test set. The UNICAMP team also has reported near-perfect results with one false acceptance and one false rejection on the development set, and one false acceptance on the test set. In spite of good performance on the development set, the SIANI team's detection method is not able to generalize the classification on the test set. Table 2 presents the usage of techniques by different teams for detecting attacks.

Team	Development		Test		
	FAR	FRR	FAR	FRR	HTER
AMILAB	0.00	0.00	0.00	1.25	0.63
CASIA	1.67	1.67	0.00	0.00	0.00
IDIAP	0.00	0.00	0.00	0.00	0.00
SIANI	1.67	1.67	0.00	21.25	10.63
UNICAMP	1.67	1.67	1.25	0.00	0.63
UOULU	0.00	0.00	0.00	0.00	0.00

Table 1. Performance figures of the different teams. All values are in percentage (%).

Team	Motion Analysis	Texture Analysis	Liveness Detection
AMILAB	√	√	√
CASIA	√	√	·
IDIAP	·	√	·
SIANI	√	·	·
UNICAMP	√	√	√
UOULU	·	√	·

Table 2. Different teams on the usage of techniques. Here, √ means the team is using corresponding technique.

5. Conclusion

To the best of our knowledge this is the first competition to compare state-of-the-art counter measures to 2-D facial spoofing attacks on a publicly available database. From the results, we clearly see that motion analysis, texture analysis, and liveness detection are three important means to obtain the clues for detecting print based spoof attacks. The usage of one or multiple techniques for detection appears to be a common trend. However, the usage of a single technique also has shown to be efficient.

A possible future investigation would be to compute performance at regular intervals of time instead of obtaining score for the whole video at once. The majority of the teams are able to clearly separate attacks from real accesses. This suggests that the problem should be made more complex, for instance expanding the database with photo quality print attacks.

6. Acknowledgments

Authors would like to thank Swiss Innovation Agency (CTI Project Replay) and FP7 European TABULA RASA Project, http://www.tabularasa-

euproject.org (257289) for the financial support. Authors would also like to thank Christine Marcel, and Flavio Tarsetti for building the acquisition system to make the database. UNICAMP's 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] J. Bai, T. Ng, X. Gao, and Y. Shi. Is Physics-based Liveness Detection Truly Possible with a Single Image? In *IEEE Intl. Symposium on Circuits and Systems*, pages 3425–3428, 2010.

[2] W. Bao, H. Li, N. Li, and W. Jiang. A Liveness Detection Method for Face Recognition based on Optical Flow Field. In *IEEE Intl. Conference on Image Analysis and Signal Processing*, pages 233–236, 2009.

[3] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. ENCARA2: Real-time Detection of Multiple Faces at Different Resolutions in Video Streams. *Journal of Visual Communication and Image Representation*, 18(2):130–140, 2007.

[4] M. Chau and M. Betke. Real time Eye Tracking and Blink Detection with USB Cameras. Technical report, 2005.

[5] C. Chow and C.N.Liu. Approximating Discrete Probability Distributions with Dependence Trees. *IEEE Trans. on Information Theory*, 14:426–467, 1968.

[6] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[7] D. Donoho. De-noising by Soft-thresholding. *IEEE Trans. on Information Theory*, 41(3):613–627, 1995.

[8] D. Donoho and I. Johnstone. Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika*, 81(3):425–455, 1994.

[9] A. H. Eid, M. N. Ahmed, B. E. Cooper, and E. E. Rippetoe. Characterization of Electrophotographic Print Artifacts: Banding, Jitter, and Ghosting. *IEEE Trans. on Image Processing*, 20:1313–1326, 2011.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1), 2009.

[11] R. Haralick, K. Shanmugam, and I. Dinstein. Texture Features for Image Classification. *IEEE Trans. on Systems, Man, and Cybernetics*, 3(6), 1973.

[12] M. Heikkila and M. Pietikainen. A Texture-based Method for Modeling the Background and Detecting Moving Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(4):657–662, 2006.

[13] K. Kollreider, H. Fronthaler, and J. Bigun. Evaluating Liveness by Face Images and the Structure Tensor. pages 75–80, 2005.

[14] K. Kollreider, H. Fronthaler, and J. Bigun. Verifying Liveness by Multiple Experts in Face Biometrics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2008.

[15] J. Li, Y. Wang, T. Tan, and A. K. Jain. Live Face Detection based on the Analysis of Fourier Spectra. In *SPIE Proceedings on Biometric Technology for Human Identification*, pages 296–303, 2004.

[16] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian. Foreground Object Detection from Videos Containing Complex Background. In *ACM Intl. Conference on Multimedia*, pages 2–10, 2003.

[17] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24:971–987, 2002.

[18] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera. *IEEE Intl. Conference on Computer Vision*, pages 1–8, 2007.

[19] G. Pan, Z. Wu, and L. Sun. Liveness Detection for Face Recognition. *Recent Advances in Face Recognition*, pages 109–124, 2008.

[20] J. Pearl and S. Russell. *Handbook of Brain Theory and Neural Networks*, chapter Bayesian Networks. Cambridge, 2002.

[21] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 763–770, 2010.

[22] W. R. Schwartz, R. D. da Silva, L. S. Davis, and H. Pedrini. A Novel Feature Descriptor Based on the Shearlet Transform. In *IEEE Intl. Conference on Image Processing*, 2011. To appear.

[23] W. R. Schwartz, H. Guo, and L. S. Davis. A Robust and Scalable Approach to Face Identification. In *European Conference on Computer Vision*, volume 5, pages 476–489, 2010.

[24] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In *IEEE Intl. Conference on Computer Vision*, 2009.

[25] R. Tronci, G. Giacinto, and F. Roli. Dynamic score combination: A supervised and unsupervised score combination method. In *Machine Learning and Data Mining in Pattern Recognition*, volume 5632 of *Lecture Notes in Computer Science*, pages 163–177. Springer Berlin / Heidelberg, 2009.

[26] P. Viola and M. Jones. Robust real-time face detection. *Intl. Journal of Computer Vision*, 57(2):137–154, 2004.

[27] 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.