$LBP-TOP$ based countermeasure against face spoofing attacks

Tiago de Freitas Pereira[1], André Anjos[2], José Mario De Martino[1], Sébastien Marcel[2]

[1]School of Electrical and Computer Engineering - University of Campinas (UNICAMP)
[2]IDIAP Research Institute tiagofrepereira@gmail.com, andre.anjos@idiap.ch, martino@fee.unicamp.br, marcel@idiap.ch

Abstract. User authentication is an important step to protect information and in this field face biometrics is advantageous. Face biometrics is natural, easy to use and less human-invasive. Unfortunately, recent work has revealed that face biometrics is vulnerable to spoofing attacks using low-tech cheap equipments. This article presents a countermeasure against such attacks based on the $LBP-TOP$ operator combining both space and time information into a single multiresolution texture descriptor. Experiments carried out with the REPLAY ATTACK database show a Half Total Error Rate ($HTER$) improvement from 15.16% to 7.60%.

1 Introduction

Despite the progress in the last years, automatic face recognition is still an active research area. Many tasks, such as recognition under occlusion or recognition in a crowd and with complex illumination conditions still represent unsolved challenges. Advances in the area were extensively reported in [8] and [16]. However, the issue of verifying if the face presented to a camera is indeed a face from a real person and not an attempt to deceive (spoof) the system has received less attention.

A spoofing attack consists in the use of forged biometric traits to gain illegitimate access to secured resources protected by a biometric authentication system. The lack of resistance to attacks is not exclusive to face biometrics. [23], [14] and [18] indicate that fingerprint authentication systems suffer from similar weakness. [11], [12] and [19] diagnose the same shortcoming on iris recognition systems. Finally, [5] and [7] address spoofing attacks to speaker biometrics. The literature review for spoofing in face recognition systems will be presented in Section 2.

In authentication systems based on face biometrics, spoofing attacks are usually perpetrated using photographs, videos or forged masks. Moreover, with the increasing popularity of social networks websites (facebook, flicker, youtube, instagram and others) a great deal of multimedia content is available on the web that can be used to spoof a face authentication system. In order to mitigate the

vulnerability of face authentication systems, effective countermeasures against face spoofing have must be deployed.

In this context, we proposed a novel countermeasure against face spoofing. Our approach uses an operator called Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) that combines space and time information into a single descriptor with a multiresolution strategy. Experiments conducted using the REPLAY ATTACK database [6] indicate that our approach has a better performance in detecting face spoofing attacks using photographs and videos than state-of-the-art techniques.

The remainder of the paper is organized as follows: Section 2 briefly review the relevant literature. Section 3 discusses the application of Local Binary Patterns (LBP) in space and time domains. Section 4 presents our approach against facial spoofing attacks. Our experimental set-up and results are discussed in Section 5. Finally, in Section 6 we summarize this work highlighting its main contributions.

2 Prior work

Considering the type of countermeasures that do not require user collaboration, Chakka et al. in [4] made a classification considering the following cues in spoofing attacks:

- Presence of vitality (liveness);
- Differences in motion patterns;
- Differences in image quality assessment.

Presence of vitality or **liveness** detection consists in the search of features that only live faces can possess. For example, Pan et al. in [20] develop a countermeasure based on eye-blink.

The countermeasures based on differences in **motion** patterns rely on the fact that real faces displays different motion behavior compared to a spoof attempt. Kollreider et al. [13] present a motion based countermeasure that estimates the correlation between different regions of the face using optical flow. In that countermeasure, the input is considered a spoof if the optical flow field on the center of the face and on the center of the ears present the same direction. The performance was evaluated using the subset "Head Rotation Shot" of the XM2VTS database whose real access was the videos of this subset and the attacks were generated with hard copies of those data. With this database, that was not made publicly available, an Equal Error Rate (EER) of 0.5% was achieved. Anjos et al. [3] present a motion based countermeasure measuring the correlation between the face and the background through simple frame differences. With the PRINT ATTACK database, that approach presented a good discrimination power ($HTER$ equals to 9%).

Countermeasures based on differences in **image quality assessment** rely on the presence of artifacts intrinsically present at the attack media. Such remarkable properties can be originated from media quality issues or differences

in reflectance properties. Li et al. [15] hypothesize that fraudulent photographs have less high frequency components than real ones. To test the hypothesis a small database was built with 4 identities containing both real access and printed photo attacks. With this **private** database, an accuracy of 100% was achieved. Because of differences in reflectance properties, real faces very likely present different texture patterns compared with fake faces. Following that hypothesis, Maatta et al. [17] and Chingovska et al. [6] explored the power of Local Binary Patterns (LBP) as a countermeasure. Maatta et al. combined 3 different LBP configurations ($LBP^{u2}_{8,2}$, $LBP^{u2}_{16,2}$ and $LBP^{u2}_{8,1}$) in a normalized face image and trained a SVM classifier to discriminate real and fake faces. Evaluations carried out with NUAA Photograph Impostor Database [21] showed a good discrimination power (2.9% in EER). Chingovska et al. analyzed the effectiveness of $LBP^{u2}_{8,1}$ and set of extended LBPs [22] in still images to discriminate real and fake faces. Evaluations carried out with three different databases, the NUAA Photograph Impostor Database, REPLAY ATTACK database and CASIA - Face Anti-spoofing Database [24] showed a good discrimination power with $HTER$ equals to 15.16%, 19.03% and 18.17% respectively. Assuming that real access images concentrate more information in a specific frequency band, Zhang et al. [24] used, as countermeasure, a set of DoG filters to select a specific frequency band to discriminate attacks and non attacks. Evaluations carried out with the CASIA - Face Anti-spoofing Database showed an Equal Error Rate of 17.00%.

3 *LBP* in space and time domain

Maatta et al. [17] and Chingovska et al. [6] propose LBP based countermeasures to spoofing attacks based on the hypothesis that real faces present different texture patterns in comparison with fake ones. However, the proposed techniques analyze each frame in isolation, not considering the behavior over time. As pointed out in Section 2, motion is a cue widely used and in combination with texture can generate a powerful countermeasure.

The first attempt to extend LBP to image sequences, exploring the space and time information, was introduced with the concept of Volume Local Binary Patterns ($VLBP$) [25]. To capture interframe patterns in textures, $VLBP$ considers the frame sequence as a parallel sequence. Considering a 3×3 kernel and thresholding the surroundings of each pixel with the central pixel of the frame sequence, the result is considered a binary value and its decimal representation is:

$$VLBP_{L,P,R} = \sum_{q=0}^{3P+1} f(i_c - i_q)2^q, \tag{1}$$

where L corresponds to the number of predecessors and successors frames, P is the number of neighbors of i_c that corresponds to the gray intensity of the evaluated pixel, i_q corresponds to the gray intensity of a specific neighbor of i_c, R is the radius of considered neighborhood and $f(x)$ is defined as follows:

$$f(x) = \begin{cases} 0 \; if \; x < 0 \\ 1 \; if \; x \geq 0 \end{cases}. \tag{2}$$

An histogram of this descriptor, contains 2^{3P+1} elements. Considering $P = 8$ (the most common configuration [6] [17] [1]) the number of bins in such histogram will be $33,554,432$ which is not computationally tractable.

To address this issue, [25] presented a simplification of the $VLBP$ operator; the so called LBP from Three Orthogonal Planes ($LBP - TOP$). Instead of considering the frame sequence as a three parallel planes, the $LBP - TOP$ consider three orthogonal planes intersecting the center of a pixel in the XY direction (normal LBP [1]), XT direction and YT direction, where T is the time axis (the frame sequence). Considering three orthogonal planes intersecting each pixel in a frame sequence, three different histograms are generated and then concatenated, as it can be seen in Fig. 1. With this approach, the size of the histogram decreases to $3 * 2^P$.

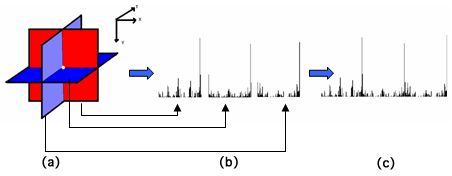

Fig. 1. (a) Three planes intersecting one pixel (b) LBP histogram of each plane (c) Concatenating the histograms (courtesy of [25]).

In the $LBP - TOP$ representation, the radii in each direction (R_X, R_Y and R_T) and the number of sampling points in each plane (P_{XY}, P_{XT} and P_{YT}) can be different as well as the type of LBP operator in each plane. They can follow the normal, the uniform pattern ($u2$) or rotation invariant uniform pattern ($riu2$) approaches [10], for example. The representation of the $LBP - TOP$ descriptor is denoted as $LBP{-}TOP^{operator}_{P_{XY},P_{XT},P_{YT},R_X,R_Y,R_T}$. In addition to the computational simplification, compared with $VLBP$, $LBP - TOP$ has the advantage to generate independent histograms for each of intersecting planes, in space and time, which can be treated in combination or individually.

Because of the aforementioned complexity issues on the implementation of a $VLBP$ based processor, the developed countermeasure uses $LBP - TOP$ to extract spatio-temporal information from video sequences.

4 The proposed countermeasure

Fig. 2 shows a block diagram of the proposed countermeasure. First, each frame of the *original frame sequence* was gray-scaled and passed through a face detector

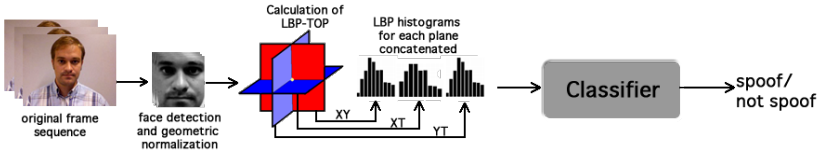

Fig. 2. Block diagram of the proposed countermeasure.

using MCT features [9]. Only *detected faces* with more than 50 pixels of width and height were considered. The detected faces were geometric normalized to 64×64 pixels. In order to reduce the face detector noise, for each set of frames used in the $LBP-TOP$ calculation, the same face bounding box was used. As can be seen in the Fig. 3, the middle frame was chosen. Unfortunately, the face detector is not error free and in case of error in the middle frame face detection, the nearest detection was chosen otherwise the observation was discarded.

After face detection step, the LBP *operators* were calculated for each plane (XY, XT and YT) and the *histograms* were computed and then concatenated.

To generate a multiresolution description, the histograms in time domain (XT and YT) are concatenated for different values of R_t. The notation chosen to represent these settings is using brackets for the multiresolution data. For example, $R_t = [1, 3]$ means that the $LBP-TOP$ operator will be calculated for $R_t = 1$, $R_t = 2$ and $R_t = 3$ and all resultant histograms will be concatenated. After the feature extraction step, this data is ready for binary *classification* to discriminate spoofing attacks from real accesses.

In order to be comparable with [6], each observation in the original frame sequence will generate a score independent of the rest of the frame sequence.

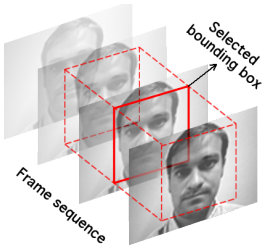

Fig. 3. Face detection strategy for $R_t = 1$.

The proposed countermeasure was implemented using the free signal-processing and machine learning toolbox Bob [2] and the source code of the algorithm is available as an add-on package to this framework[1].

[1] http://pypi.python.org/pypi/antispoofing.lbptop

5 Experiments

This section describes the performance evaluation of the proposed countermeasure on the REPLAY-ATTACK database [6] and using its defined protocol. Such protocol defines 3 non-overlapped partitions for training, development and testing countermeasures. The training set should be used to train the countermeasure, the development set is used to tune the parameters of the countermeasure and to estimate a threshold value to be used in the test set. The protocol defines the Equal Error Rate (EER) as a decision threshold. Finally, the test set must be used only to report results. As performance measurement, the protocol suggests to report the Half Total Error Rate ($HTER$) on the test data.

5.1 Evaluation methodology

In order to measure the effectiveness of this countermeasure, each parameter was tuned solely (fixing other elements) using the development set. For this, 5 experiments were carried out evaluating the effectiveness of:

1. Each $LBP-TOP$ plane;
2. Different classifiers;
3. Different LBP operators;
4. Different numbers of sampling points in the $LBP-TOP$ operator
5. Multiresolution approach.

Inspired on [6], the $LBP-TOP$ operator chosen to start the evaluation was $LBP-TOP^{u2}_{8,8,8,1,1,R_T}$.

5.2 Effectiveness of each $LBP-TOP$ plane

Fig. 4 shows the evolution of the test set $HTER$ considering individual and combined histograms of $LBP-TOP$ planes. First, it was analyzed the effectiveness of each individual plane and then combinations when the multiresolution area (R_t) is increased. We used, as binary classifier, a linear projection derived from Linear Discriminant Analysis LDA as is [6].

It can be seen that, by combining the time components (XT and YT planes) the results were improved. This suggests that the time information is an important cue. The combination of the three planes generated the best results which suggests that both spatial and time information are important to classify real and fake faces. For that reason, next results will be presented always with a combination of the three $LBP-TOP$ planes (XY, XT and YT).

5.3 Effectiveness of different classifiers

Fig. 5 shows the performance of this countermeasure, in $HTER$ terms, with different classifiers when the multiresolution area (R_t) is increased. The first classifier applied was the χ^2 distance, since the feature vectors are histograms.

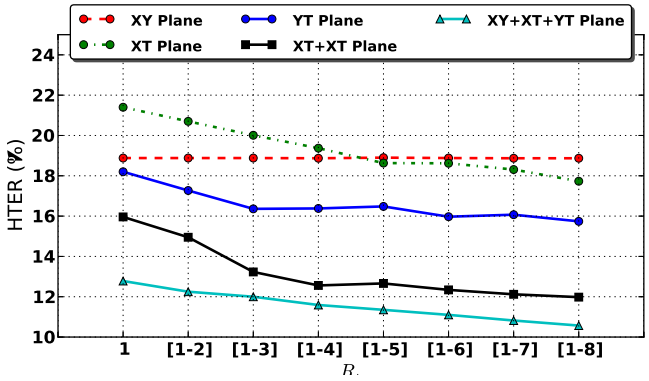

Fig. 4. (Color online) Evaluation of HTER(%) in each plane when the multiresolution area (R_t) is increased with $LBP-TOP^{u2}_{8,8,8,1,1,R_t}$ and LDA classifier - test-set.

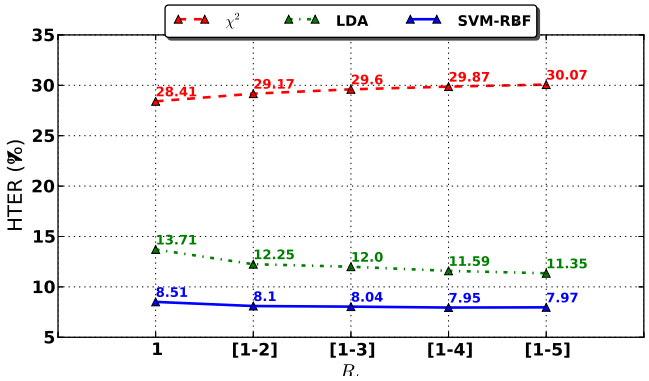

Fig. 5. (Color online) Evaluation of $HTER(\%)$ with $LBP-TOP^{u2}_{8,8,8,1,1,R_t}$ using different classifiers.

For that, the same strategy adopted in [6] was carried out. A reference histogram only with real accesses was created averaging the histograms in the training set. Experiments using more complex classifiers were carried out as well. For that, Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) with a radial basis function kernel (RBF) were chosen.

It can be seen that best results were obtained with the non linear SVM using RBF kernel. It is important to remark that results presented with SVM, should be analyzed carefully for overtraining. The final machine uses $\sim$ 25000 support vectors to achieve 7.97%. This number represents $\sim$ 33% of the training set size. A simple comparison with the same $LBP-TOP$ configuration with LDA classifier resulted in an $HTER$ equal to 11.35%. This is not a huge gap and the classifier is far simpler.

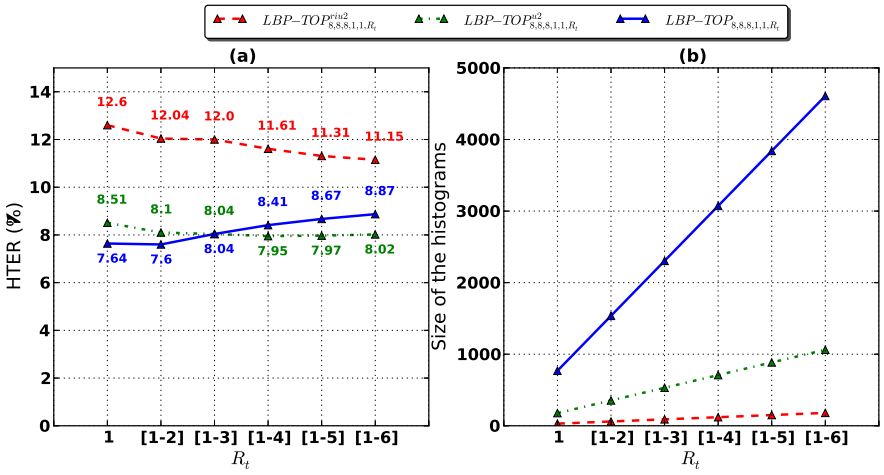

Fig. 6. (Color online) (a) Evaluation of $HTER(\%)$ with $LBP - TOP_{8,8,8,1,1,R_t}$ using different LBP configurations in the planes with SVM classifier (b) Evaluation of the histogram size when (R_t) is increased.

5.4 Effectiveness of different LBP operators

The size of the histogram in a multiresolution analysis, in time domain, increases linearly with R_t. The choice of an appropriate LBP representation in the planes is an important issue since this choice impacts the size of the histograms. Using uniform patterns or rotation invariant extensions, in one or multiple planes, may bring a significative advantage in computational complexity. Fig. 6 (a) shows the performance, in $HTER$ terms, configuring each plane as normal LBP (with 256 bins for $P = 8$), LBP^{u2} and LBP^{riu2} when the multiresolution area (R_t) is increased. Results must be interpreted with the support with the Fig. 6 (b), which shows the number of bins on the histograms used for classifications in each configuration.

It can be seen that, when R_t is increased, the $HTER$ saturates in $\sim 11\%$ and $\sim 8\%$ for LBP^{riu2} and LBP^{u2} respectivelly. The normal LBP operator presents a minimum in 7.60% with $R_t = [1, 2]$ (the best result achieved in this paper). Results with LBP and LBP^{u2} presented similar performance and even the LBP presented the best result, using LBP^{u2} seems a reasonable tradeoff between computational complexity and performance (in $HTER$ terms). Hence we will still proceed with LBP^{u2}.

5.5 Effectiveness of different numbers of sampling points in the $LBP-TOP$ operator

Another parameter that impacts in the size of the histograms is the number of sampling points (P) in each plane. Fig. 7(a) and (b) show the performance, in $HTER$, varying the values of P_{XT} and P_{YT} to 4, 8 and 16 when the multiresolution area (R_t) is increased with SVM and LDA classifiers respectively.

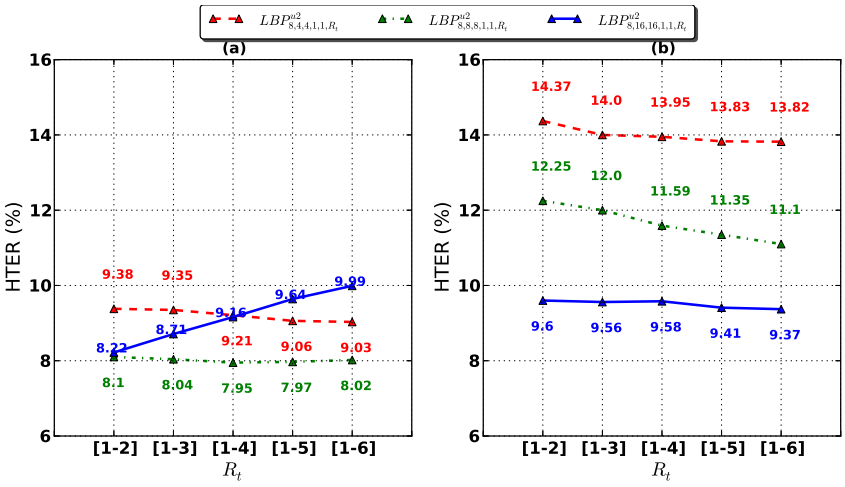

Fig. 7. (Color online) Evaluation of $HTER(\%)$ with $LBP-TOP^{u2}$ using different values for P_{XT} and P_{YT} in the time planes using **(a)** SVM classifier **(b)** LDA classifier.

It can be seen that results with $LBP-TOP^{u2}_{8,8,8,1,1,R_t}$ achieved the best performance (saturating around 8%), using an SVM classifier (see Fig. 7(a)). However, it was expected good performance using P_{XT} and P_{YT} set to 16 when the multiresolution analysis (R_t) is increased, since more points were extracted over the time. Observing the Fig. 7 (b), with LDA as a classifier, the best performance was achieved with P_{XT} and P_{YT} equal to 16. These results suggests that, when the multiresolution area is increased with P_{XT} and P_{YT} equals to 16, the SVM classifier loses generalization power. In order to track that hypothesis, a simple observation in the number of support vectors can be done. Not surprisingly, the number of support vectors increases from ~ 30000 to ~ 35000 for R_t equals to $[1,2]$ and $[1,6]$ respectively. That increase, in the final SVM, represents $\sim 32\%$ and $\sim 39\%$ of the training set size respectively, re-assign the overtraining hypothesis. Hence we will still proceed with $LBP^{u2}_{8,8,8,1,1,R_t}$ for the next experiment.

5.6 Effectiveness of multiresolution approach

Fig. 8 shows the performance of this countermeaure considering a multiresolution approach compared with a single resolution approach. The single resolution

approach consists in using only fixed values for R_t, without concatenating histograms for each R_t. With this approach the size of the histograms will be constant along R_t increase, what decreases the computational complexity.

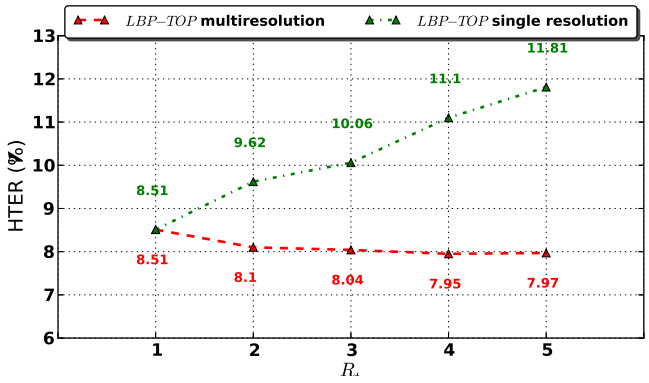

Fig. 8. (Color online) Evalutation of $HTER\%$ using $LBP-TOP^{u2}_{8,8,8,1,1,R_t}$ with and without histogram concatenation using SVM classifier.

It can be seen that, when the single resolution approach is considered, the HTER increases with R_t whereas the multiresolution approach helps to keep the HTER low with the increasing value of R_t. It is possible to suggest that, for the $LBP-TOP$ descriptor, motion patterns between closest frames carry more information for spoofing detection than distant ones. Nevertheless, information from distant frames are important as well and thats help to explain why the best results were achieved with the multiresolution approach.

5.7 Summary

Table 5.7 summarizes all results obtained compared with the state of art results. The two first rows are results presented in [6] and the third row was a countermeasure based on [3] whose source code is freely available for comparison. It can be seen that the proposed countermeasure presented the best results, overtaking the state of art results in the REPLAY ATTACK database.

6 Conclusion

This article presented a countermeasure against face spoofing attacks using the $LBP-TOP$ descriptor combining both space and time information into a single descriptor. Experiments carried out with the REPLAY ATTACK database showed that an analysis in time domain improved the results comparing to the still frame analysis presented in [6] and [17]. A multiresolution analysis in time domain shows even better results, achieving 7.60% when combined with an SVM classifier (the best result achieved). It is important to remark that results with

Table 1. HTER(%) of classification with different countermeasures

	$HTER(\%)$	
	dev	**test**
$LBP^{u2}_{8,1} + SVM$ [6]	14.84	15.16
($LBP^{u2}_{8,2}$+$LBP^{u2}_{16,2}$+ $LBP^{u2}_{8,1}$) + SVM [6]	13.90	13.87
Motion coefficient based [3]	11.78	11.79
$LBP-TOP^{riu2}_{8,8,8,1,1,[1-6]} + SVM$	9.78	11.15
$LBP-TOP^{u2}_{8,4,4,1,1,[1-6]} + SVM$	8.49	9.03
$LBP-TOP^{u2}_{8,8,8,1,1,[1-4]} + SVM$	8.49	7.95
$LBP-TOP_{8,8,8,1,1,[1-2]} + SVM$	7.88	**7.60**
$LBP-TOP_{8,16,16,1,1,[1-2]} + SVM$	9.16	8.22

SVM classifier should be taken with care because with the increase of the multiresolution area, the SVM classifier tends to overtrain on the data. However, experiments with simpler classifiers, such as LDA, showed that the $LBP-TOP$ multiresolution approach still demonstrated a great potential against face spoofing in different kind of attacks scenarios, beating the state of art results.

Acknowledgement. The authors would like to thank the Swiss Innovation Agency (CTI Project Replay), the FP7 European TABULA RASA Project (257289), FUNTELL (Brazilian Telecommunication Technological Development Fund) and CPqD Telecom and IT Solutions for their financial support.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Inteligence **28** (2006) 2037–2041
2. A. Anjos, L. El Shafey, R. Wallace, M. Günther, C. McCool, S. Marcel Bob: a free signal processing and machine learning toolbox for researchers. 20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan (2012)
3. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA. (2011)
4. Chakka, M., Anjos, A., Marcel, S., Tronci, R., Muntoni, D., Fadda, G., Pili, M., Sirena, N., Murgia, G., Ristori, M., Roli, F., Yan, J., Yi, D., Lei, Z., Zhang, Z., Z.Li, S., Schwartz, W., Rocha, A., Pedrini, H., Lorenzo-Navarro, J., Castrillón-Santana, M., Maatta, J., Hadid, A., Pietikainen, M.: Competition on counter measures to 2-d facial spoofing attacks. IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA. (2011)
5. Chetty, G., Wagner, M.: Liveness verification in audio-video speaker authentication. In: Proceeding of International Conference on Spoken Language Processing ICSLP. Volume 4. (2004) 2509–2512
6. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. IEEE BIOSIG 2012

7. Eveno, N., Besacier, L.: A speaker independent" liveness" test for audio-visual biometrics. In: Ninth European Conference on Speech Communication and Technology. (2005)
8. Flynn, P., Jain, A., Ross, A.: Handbook of biometrics. Springer (2008)
9. Froba, B., Ernst, A.: Face detection with the modified census transform. In: Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, IEEE (2004) 91–96
10. Inen, M., Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. Volume 40. Springer Verlag (2011)
11. Johnson, P., Tan, B., Schuckers, S.: Multimodal fusion vulnerability to non-zero effort (spoof) imposters. In: Information Forensics and Security (WIFS), 2010 IEEE International Workshop on, IEEE (2010) 1–5
12. Kanematsu, M., Takano, H., Nakamura, K.: Highly reliable liveness detection method for iris recognition. In: SICE, 2007 Annual Conference, IEEE (2007) 361–364
13. Kollreider, K., Fronthaler, H., Bigun, J.: Non-intrusive liveness detection by face images. Image and Vision Computing **27** (2009) 233–244
14. Leyden, J.: Gummi bears defeat fingerprint sensors. The Register - **16** 2002
15. Li, J., Wang, Y., Tan, T., Jain, A.: Live face detection based on the analysis of fourier spectra. Biometric Technology for Human Identification **5404** (2004) 296–303
16. Li, S., Jain, A.: Handbook of face recognition. Springer (2011)
17. Maatta and, J., Hadid, A., Pietika andinen, M.: Face spoofing detection from single images using texture and local shape analysis. Biometrics, IET **1** (2012) 3 –10
18. Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S.: Impact of artificial gummy fingers on fingerprint systems. In: Proceedings of SPIE. Volume 4677. (2002) 275–289
19. Pacut, A., Czajka, A.: Aliveness detection for iris biometrics. In: Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International, IEEE (2006) 122–129
20. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on Computer Vision, IEEE (2007) 1–8
21. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. Computer Vision–ECCV 2010 (2010) 504–517
22. Trefnỳ, J., Matas, J.: Extended set of local binary patterns for rapid object detection. In: Proceedings of the Computer Vision Winter Workshop. Volume 2010. (2010)
23. Uludag, U., Jain, A.: Attacks on biometric systems: a case study in fingerprints. In: Proc. SPIE-EI. (2004) 622–633
24. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.: A face antispoofing database with diverse attacks. In: Biometrics (ICB), 2012 5th IAPR International Conference on Biometrics, IEEE (2012) 26–31
25. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29** (2007) 915 –928