Face Spoofing Detection Using Dynamic Texture

Jukka Komulainen, Abdenour Hadid, and Matti Pietikäinen

Center for Machine Vision Research,
Department of Computer Science and Engineering,
P. O. Box 4500 FI-90014 University of Oulu, Finland
{jukmaatt,hadid,mkp}@ee.oulu.fi

Abstract. While there is a significant number of works addressing e.g. pose and illumination variation problems in face recognition, the vulnerabilities to spoofing attacks were mostly unexplored until very recently when an increasing attention is started to be paid to this threat. A spoofing attack occurs when a person tries to masquerade as someone else e.g. by wearing a mask to gain illegitimate access and advantages. This work provides the first investigation in research literature on the use of dynamic texture for face spoofing detection. Unlike masks and 3D head models, real faces are indeed non-rigid objects with contractions of facial muscles which result in temporally deformed facial features such as eye lids and lips. Our key idea is to learn the structure and the dynamics of the facial micro-textures that characterise only real faces but not fake ones. Hence, we introduce a novel and appealing approach to face spoofing detection using the spatiotemporal (dynamic texture) extensions of the highly popular local binary pattern approach. We experiment with two publicly available databases consisting of several fake face attacks of different natures under varying conditions and imaging qualities. The experiments show excellent results beyond the state-of-the-art.

1 Introduction

Because of its natural and non-intrusive interaction, identity verification and recognition using facial information is among the most active and challenging areas in computer vision research. Despite the significant progress in the face recognition technology in the recent decades, wide range of viewpoints, aging of subjects and complex outdoor lighting are still research challenges. While there is a significant number of works addressing these issues, research on face biometric systems under spoofing attacks has mostly been overlooked although face recognition systems are known, since long time ago, to respond weakly to attacks. A spoofing attack occurs when a person tries to masquerade as someone else by falsifying data and thereby gaining illegitimate access. Very recently, an increasing attention is started to be paid to the problem of spoofing attacks against face biometric systems. This can be attested by the recently organized IJCB 2011 competition on counter measures to 2D facial spoofing attacks [1] which can be seen as a kick-off for studying best practices for non-intrusive spoofing detection.

J.-I. Park and J. Kim (Eds.): ACCV 2012 Workshops, Part I, LNCS 7728, pp. 146–157, 2013.

One can spoof a face recognition system by presenting a photograph, a video or a 3D model of a targeted person to the camera. While one can also use make-up or plastic surgery as other means of spoofing, photographs are probably the most common sources of spoofing attacks because one can easily download and capture facial images. Typical countermeasure against spoofing is liveness detection that aims at detecting physiological signs of life such as eye blinking, facial expression changes, mouth movements, etc. For instance, Pan *et al.* [2] exploited the observation that humans blink once every 2-4 seconds and proposed an eye blink-based anti-spoofing method. Another commonly used countermeasure is motion analysis since it can be assumed that the movement of planar objects (e.g. displays and photographs) differs significantly from that of real human faces which are complex non-rigid 3D objects [3,4]. Obviously, such countermeasures can only be considered with photographs while nowadays videos are ubiquitous and hence can easily be used for spoofing attacks. Another category of anti-spoofing methods are based on the analysis of skin properties such as skin texture and skin reflectance. An intuitive approach is to explore the high frequency information in the facial region, since mobile phone displays and smaller photographs probably contain fewer high frequency components compared to real faces [5,6]. Such an approach is likely to fail with higher quality photographs and videos, as shown for example in [7]. Recently, also micro-texture analysis has been applied to measure facial texture quality with impressive results [8,9]. However, the evaluations were made using data sets with little variations and the used high frequency information depends strongly on the input image and fake face quality. Other countermeasures against face spoofing attacks include multi-modal analysis and multi-spectral methods. A system combining face recognition with other biometric modalities such as gait and speech is indeed intrinsically more difficult to spoof than uni-modal systems. Multi-spectral imaging can also be used for analyzing the reflectance of object surfaces and thus discriminating live faces from fake ones [10].

It appears that most of the existing methods for spoofing detection are either very complex (and hence not very practical for real-world face biometric systems requiring fast processing) or using non-conventional imaging systems (e.g. multi spectral imaging) and devices (e.g. thermal cameras). We therefore propose in this work a novel computationally fast approach based on highly discriminative dynamic micro-texture features, using conventional images and requiring no user-cooperation.

This work provides **the first investigation in research literature** on the use of dynamic texture for face spoofing detection. Unlike masks and 3D head models, real faces are indeed non-rigid objects with contractions of facial muscles which result in temporally deformed facial features such as eye lids and lips. Our key idea is to learn the structure **and** especially the **dynamics** of the facial micro-textures that characterize only real faces but not fake ones. Hence, we introduce the first and appealing spatio-temporal approach to face spoofing detection using the spatiotemporal (dynamic texture) extensions of the highly popular local binary pattern (LBP) approach [11]. Spatiotemporal LBP has shown very promising performance in various problems, including dynamic texture recognition, face and facial expression recognition, lip-reading, and activity and gait recognition [11].

Dynamic textures provide a new and very effective tool for motion analysis. The past research on motion analysis has been usually based on assumption that the scene is Lambertian, rigid and static. For example, the Lambertian assumption has been crucial when developing methods for tracking, determining optical flow or finding correspondences. This kind of constraints greatly limits the applicability of motion analysis. Recently, approaches based on dynamic textures have been proposed as a new and potentially very effective tool for motion analysis [11]. These developments have inspired us to approach face spoofing detection from dynamic texture point of view. We introduce below our novel approach and provide extensive experimental analysis on two publicly available databases (CASIA Face Anti-Spoofing Database [7] and Print-Attack Database [12]) consisting of several fake face attacks of different natures and under varying conditions and imaging qualities, showing excellent results beyond the state-of-the-art.

2 Spatiotemporal Face Liveness Description

For describing the face liveness for spoofing detection, we considered an elegant approach to face analysis from videos which is based on a spatiotemporal representation for combining facial appearance and dynamics. We adopted the LBP based spatiotemporal representation because of its recent excellent performance in modeling moving faces for face and facial expression recognition and also for dynamic texture recognition. More specifically, we considered local binary patterns from three orthogonal planes (LBP-TOP) which have shown to be very effective in describing the horizontal and vertical motion patterns in addition to appearance [13].

The LBP texture analysis operator, introduced by Ojala *et al.* [14,15], is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful texture descriptor and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes. The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood with the center value and considering the result as a binary number. The histogram of these $2^8 = 256$ different labels can then be used as a image descriptor.

The original LBP operator was defined to only deal with the spatial information. Recently, it has been extended to a spatiotemporal representation for dynamic texture analysis (DT). This has yielded to so called Volume Local Binary Pattern operator (VLBP) [13]. The idea behind VLBP consists of looking at dynamic texture as a set of volumes in the (X,Y,T) space where X and Y denote the spatial coordinates and T denotes the frame index (time). The neighborhood of each pixel is thus defined in a three dimensional space. Then, similarly to basic LBP in spatial domain, volume textons can be defined and extracted into histograms. Therefore, VLBP combines motion and appearance into a dynamic texture description.

To make the VLBP computationally simple and easy to extend, the co-occurrences of the LBP on the three orthogonal planes (LBP-TOP) was also introduced [13]. LBP-TOP consists of the three orthogonal planes: XY, XT and YT, and concatenating local binary pattern co-occurrence statistics in these three directions. The circular neighborhoods are generalized to elliptical sampling to fit to the space-time statistics. The LBP codes are extracted from the XY, XT and YT planes, which are denoted as $XY - LBP$, $XT - LBP$ and $YT - LBP$, for all pixels, and statistics of the three different planes are obtained, and then concatenated into a single histogram. The procedure is shown in Fig. 1. In this representation, dynamic texture (DT) is encoded by the $XY - LBP$, $XT - LBP$ and $YT - LBP$.

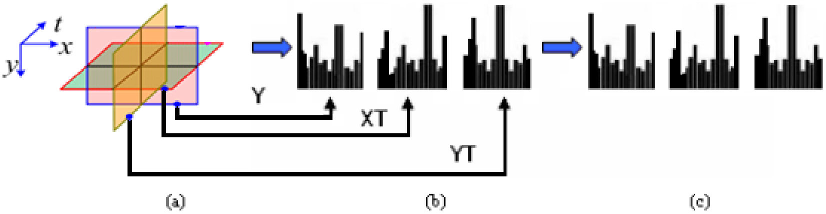

Fig. 1. (a) Three planes of dynamic texture; (b) LBP histogram from each plane; (c) Concatenated feature histogram [13]

Using equal radiuses for the time and spatial axes is not reasonable for dynamic textures [13] and therefore, in the XT and YT planes, different radii can be assigned to sample neighboring points in space and time. More generally, the radii in axes X, Y and T, and the number of neighboring points in the XY, XT and YT planes can also be different denoted by R_X, R_Y and R_T, P_{XY}, P_{XT} and P_{YT}. The corresponding feature is denoted as $LBP - TOP_{P_{XY},P_{XT},P_{YT},R_X,R_Y,R_T}$.

Let assume we are given an $X \times Y \times T$ dynamic texture ($x_c \in \{0, \cdots, X-1\}$, $y_c \in \{0, \cdots, Y-1\}, t_c \in \{0, \cdots, T-1\}$). A histogram of the DT can be defined as:

$$H_{i,j} = \sum_{x,y,t} I\left\{f_j(x,y,t) = i\right\}, \ \ i = 0, \cdots, n_j - 1; j = 0, 1, 2 \ . \tag{1}$$

in which n_j is the number of different labels produced by the LBP operator in the jth plane ($j = 0 : XY,\ 1 : XT\ and\ 2 : YT$) and $f_i(x,y,t)$ expresses the LBP code of central pixel (x,y,t) in the jth plane.

Similarly to the original LBP, the histograms must be normalized to get a coherent description for comparing the DTs:

$$N_{i,j} = \frac{H_{i,j}}{\sum_{k=0}^{n_j-1} H_{k,j}} \ . \tag{2}$$

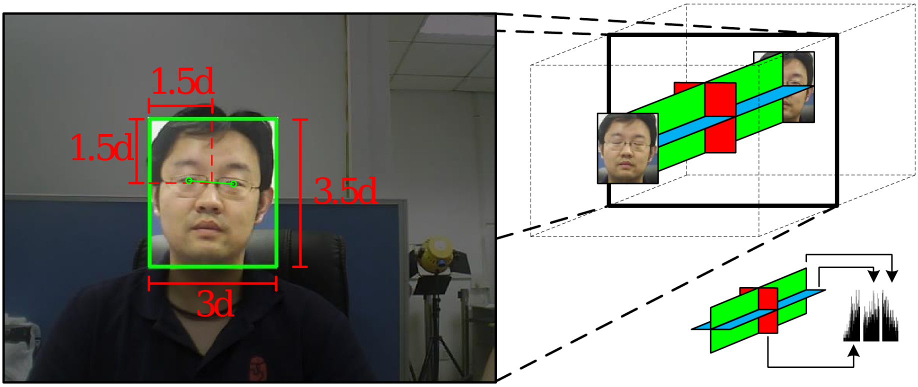

Fig. 2. Dynamic texture based face description

Due to its tolerance against monotonic gray-scale changes, LBP is adequate for measuring the facial texture quality and determining whether degradations due to spoofing medium are observed. We adopted the LBP based spoofing detection in spatiotemporal domain because LBP-TOP features have been successfully applied in describing dynamic events. Our key idea is use LBP-TOP features for detecting e.g. specific facial motion patterns or sudden characteristic reflections of planar spoofing media which might differentiate real faces from fake ones.

When deriving our proposed face liveness description, we aim to avoid scaling during geometric face normalization in order to keep all valuable information about the facial texture quality which is a crucial visual cue in spoofing detection. Simple head pose correction based on eye locations may also be too unstable between video frames, e.g. due to inaccurate eye detection, yielding performance degradation in dynamic texture analysis. To overcome these effects, LBP-TOP$_{8,8,8,1,2,2}$ operator is instead applied on each pixel and the dynamic LBP histogram for every frame is calculated over the volume bounded by the roughly normalized face location (see Fig. 2). Then, the histograms of 768 bins are accumulated over a period of two seconds (50 frames at 25fps) to form the final feature vector.

3 Experimental Analysis

To assess of the effectiveness of our proposed spatiotemporal face liveness description for spoofing detection, we performed a set of experiments on the CASIA Face Anti-Spoofing Database [7] and Print-Attack Database [12] from the Idiap Research Institute. We used Viola-Jones algorithm for face detection [16] while eye localization is performed using 2D Cascaded AdaBoost [17]. The relative eye positions from the first frame are used to refine the detected faces so that the face areas are roughly aligned in every frame. We also exploited the fact that the spoofing medium might be visible around the face, thus the height and width of the aligned face are set to 3.5d and 3.0d where d represents the distance between eyes (see Fig. 2). Once the face liveness description is derived, a homogeneous kernel map [18] is applied to obtain a five dimensional linear approximation of a χ^2 kernel. The approximated feature map is computed with VLFeat [19] and the final classification is performed using a linear SVM implementation of LIBLINEAR [20].

The SVM classifier is trained using a set of positive (genuine faces) and negative (fake faces) samples which are extracted from the provided training data. In order to get sufficient amount of data for building the model, the whole length of each training video is divided into several time windows with temporal overlap of one second over which the LBP-TOP features are computed. On the test sets, however, only the first two seconds from the beginning of each video sequence are used for determining whether a genuine face or a fake one is observed. The use of the whole video sequence may naturally lead to better detection results but at the cost of more computational time which could be an issue in real-life applications.

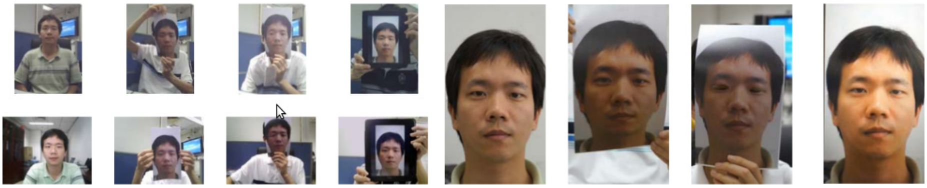

Fig. 3. Example images from the CASIA Face Anti-Spoofing Database [7]

3.1 Evaluation on the CASIA Face Anti-Spoofing Database

We first conducted extensive experiments on the CASIA Face Anti-Spoofing Database [7] and compared our results against those which are provided along with the database. The database includes significant improvements compared to previous databases, since it provides more variations in the collected data. The data set contains 50 real clients and the corresponding fake faces are captured with high quality from the original ones. The variety is achieved by introducing three imaging qualities (low, normal and high) and three fake face attacks which include warped photo, cut photo (eyeblink) and video attacks. Examples from the database can be seen in Fig. 3. Altogether the database consists of 600 video clips and the subjects are divided into subsets for training and testing (240 and 360, respectively). Results of a baseline system are also provided along the database for fair comparison. The baseline system considers the high frequency information in the facial region using multiple DoG features and SVM classifier and is inspired by the work of Tan *et al.* [6].

Since the main purpose of the database is to investigate the possible effects of different fake face types and imaging qualities, the test protocol consists of seven scenarios in which particular train and test samples are to be used. The quality test considers the three imaging qualities separately, low (1), normal (2) and high quality (3), and evaluates the overall spoofing detection perfomance under variety of attacks at the given imaging quality. Similarly, the fake face test assesses how robust the anti-spoofing measure is to specific fake face attacks, warped photo (4), cut photo (5) and video attacks (6), regardless of the imaging quality. In the overall test (7), all data is used to give a more general evaluation. The results of each scenario are reported as Detection-Error Trade-off (DET) curves and equal error rates (EER), which is the point where false acceptance rate (FAR) equals false rejection rate (FRR) on the DET curve.

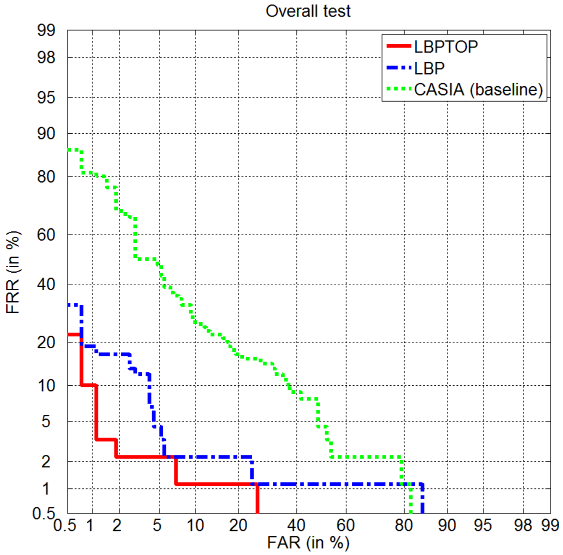

Fig. 4. Overall comparative results on the CASIA Face Anti-Spoofing Database

The results of the experiments are shown in Fig. 4 as DET curves for the overall test, i.e. including all scenarios. As it can be seen, the use of only facial appearance (LBP) leads to better results compared to the baseline method (CASIA baseline). Importantly, when the temporal planes XT and YT are also considered for spatio-temporal face description (LBP-TOP), a significant performance enhancement is obtained, thus confirming the benefits of encoding and exploiting not only the facial appearance but also the facial dynamics information.

More detailed results for each spoofing attack scenario are presented in Fig. 5 and in Table 1. The results indicate that the proposed LBP-TOP based face description yields best results in all configurations except at the highest imaging quality. The facial appearance description (LBP) works perfectly when the highest imaging quality is used because the skin texture of genuine faces looks

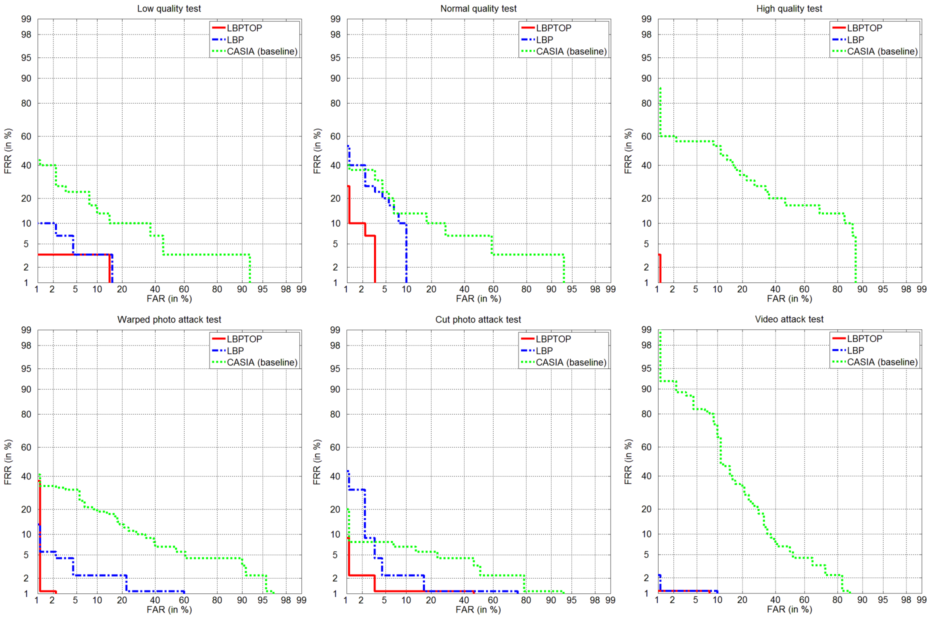

Fig. 5. DET curves under the different protocols of the CASIA Face Anti-Spoofing Database

Table 1. EER comparison between the baseline method, LBP and LBP-TOP on the CASIA Face Anti-Spoofing Database

Scenario	1	2	3	4	5	6	7
Baseline	**0.13**	**0.13**	**0.26**	**0.16**	**0.06**	**0.24**	**0.17**
LBP	**0.04**	**0.10**	**0.00**	**0.04**	**0.04**	**0.01**	**0.04**
LBPTOP	**0.03**	**0.03**	**0.01**	**0.01**	**0.02**	**0.01**	**0.02**

strikingly sharper compared to the fake ones. Thus, the measurement of facial texture quality seems to provide sufficient means to reveal whether degradation due to recapturing process is observed if the imaging quality is good enough to capture the fine details of a human face. However, the quality test shows that the use of facial dynamics enhances the spoofing detection results at lower imaging qualities without any significant performance drop when a high resolution camera is used for capturing the facial image. Furthermore, the fake face test indicates that adding temporal planes to the face description improves the robustness to different types of spoofing attacks, especially to warped and cut photo attacks, at various imaging qualities. The downsized resolution of the original high quality video spoofs (due to limited iPad screen resolution) [7] and the occasionally visible video screen frame around the fake faces also partially explain the less challening nature of the video attack tests.

3.2 Evaluation on the Print-Attack Database

For extensive evaluation, we also conducted experiments on a second publicly available database namely Print-Attack Database [12] which was originally introduced within the IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks [1]. The database consists of 200 real client accesses and 200 print-attack videos (50 clients) which were captured in controlled and uncontrolled lighting conditions using a webcam at 25fps with a resolution of 320×240 pixels. The print attacks were generated by taking high-resolution photographs of each client under the same conditions as in their authentication sessions and the captured images were printed in color on A4-sized paper. The spoofing attack attempts were performed with fixed or hand-held prints. Example images from the database are shown in Fig. 6. The database is divided into three sets, training, development and test data (see Table 2). Clients have been randomly divided for each subset so that the identities do not overlap between the subsets. The EER of development set is used for tuning the threshold which is applied for discriminating the test samples. For simplicity, we used the provided face locations for extracting the LBP-TOP representations.

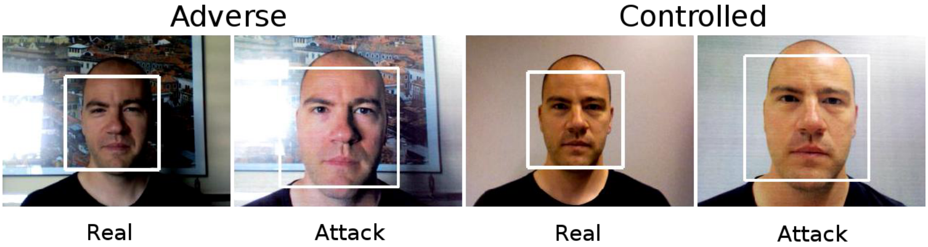

Fig. 6. Examples from Print-Attack Database [12] with the provided face locations. Note that the photo attacks suffer from apparent printing artefacts.

Table 2. The decomposition of the Print-Attack Database. The numbers indicate how many videos are included in each subset (the sums indicate the amount of hand-based and fixed-support attacks).

Type	Train	Devel.	Test	Total
Real	60	60	80	200
Attack	30+30	30+30	40+40	100+100
Total	120	120	160	400

Our dynamic texture based face description approach easily detected and characterized the printing artifacts and facial movements, e.g. eye blinking. Our approach yields perfect detection results (EER of 0%) on this database. Print attacks are perhaps less challenging to our method than the combination of different types of attacks as in the CASIA Face Anti-Spoofing Database. Table 3 shows a performance comparison between our proposed approach and the works of different research groups who participated in the IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks [1]. It is worth mentioning that our proposed dynamic texture based face description performed very well even using only a single LBP-TOP feature vector which is easily extracted from the face area, whereas the other methods considered more complex analysis using multiple cues, e.g. fusion of separate motion and texture analysis, or relying on describing the strongly visible print defects which are quite obvious in the data set.

Table 3. Performance comparison between the proposed approach and the teams who participated in the IJCB 2011 Competition on Counter Measures to 2D Facial Spoofing Attacks [1]

	Development		Test		
Method	FAR	FRR	FAR	FRR	HTER
AMILAB [1]	**0.00**	**0.00**	**0.00**	**1.25**	**0.63**
CASIA [1]	**1.67**	**1.67**	**0.00**	**0.00**	**0.00**
IDIAP [1]	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
SIANI [1]	**1.67**	**1.67**	**0.00**	**21.25**	**10.63**
UNICAMP [1]	**1.67**	**1.67**	**1.25**	**0.00**	**0.63**
UOULU [1]	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Proposed approach	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

4 Conclusion

Inspired by the recent progress in dynamic texture, we investigated the problem of face spoofing detection using spatiotemporal local binary patterns. To the best of our knowledge, this is the first work in the literature applying dynamic texture to the spoofing detection problem. The key idea of our proposed approach consists of analyzing the structure and the dynamics of the micro-textures in the facial regions using LBP-TOP features which provide an efficient and compact representation for face liveness description. Experiments on two publicly available databases showed excellent results under various fake face attacks, including video replay attacks, at different imaging qualities. The incorporation of facial dynamics significantly increased the robustness of the LBP based face description regardless of the imaging quality, especially under warped and cut photo attacks. Our obtained results can be used by the research community as a new reference on these spoofing databases for future research.

The excellent obtained results on these two publicly available databases suggest that more challenging databases (e.g. using 3D skin-like masks of very high quality and precision) should be designed, captured and made publically available for the research community in the near future. It would be then of great interest to evaluate our approach on such challenging data when available. We are currently incorporating the described anti-spoofing measure into our existing access control system for deployment in real-world applications. We plan to release the source code of our described anti-spoofing method for the research community after the publication of this work.

Acknowledgement. This work has been performed within the TABULA RASA project 7th Framework Research Programme of the European Union (EU), grant agreement number: 257289. The financial support of the Academy of Finland and Infotech Oulu Graduate School is also gratefully acknowledged.

References

1. Chakka, M.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., Li, S., Schwartz, W.R., Rocha, A., Pedrini, H., Lorenzo-Navarro, J., Castrillón-Santana, M., Määttä, J., Hadid, A., Pietikäinen, M.: Competition on counter measures to 2-d facial spoofing attacks. In: Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA (2011)
2. Pan, G., Wu, Z., Sun, L.: Liveness detection for face recognition. In: Delac, K., Grgic, M., Bartlett, M.S. (eds.) Recent Advances in Face Recognition, ch. 9. INTECH (2009)
3. Kollreider, K., Fronthaler, H., Bigun, J.: Non-intrusive liveness detection by face images. Image and Vision Computing 27, 233–244 (2009)
4. Bao, W., Li, H., Li, N., Jiang, W.: A liveness detection method for face recognition based on optical flow field. In: 2009 International Conference on Image Analysis and Signal Processing, pp. 233–236. IEEE (2009)
5. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of fourier spectra. In: Biometric Technology for Human Identification, pp. 296–303 (2004)
6. Tan, X., Li, Y., Liu, J., Jiang, L.: Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 504–517. Springer, Heidelberg (2010)
7. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: Proceedings of 5th IAPR International Conference on Biometrics (ICB 2012), New Delhi, India (2012)
8. Bai, J., Ng, T.T., Gao, X., Shi, Y.Q.: Is physics-based liveness detection truly possible with a single image? In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 3425–3428 (2010)
9. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using micro-texture analysis. In: Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA (2011)
10. Zhang, Z., Yi, D., Lei, Z., Li, S.Z.: Face liveness detection by learning multispectral reflectance distributions. In: International Conference on Face and Gesture, pp. 436–441 (2011)
11. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. Springer (2011)

12. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA (2011)
13. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 915–928 (2007)
14. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29, 51–59 (1996)
15. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. on PAMI 24 (2002)
16. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
17. Niu, Z., Shan, S., Yan, S., Chen, X., Gao, W.: 2d cascaded adaboost for eye localization. In: Proc. of the 18th International Conference on Pattern Recognition (2006)
18. Vedaldi, A., Zisserman, A.: Efficient additive kernels via explicit feature maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2010)
19. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008)
20. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research 9, 1871–1874 (2008)