Exploting Periocular and RGB Information in Fake Iris Detection

Fernando Alonso-Fernandez, Josef Bigun
Halmstad University. Box 823. SE 301-18 Halmstad, Sweden
{feralo, josef.bigun}@hh.se, http://islab.hh.se

***Abstract*—Fake iris detection has been studied by several researchers. However, to date, the experimental setup has been limited to near-infrared (NIR) sensors, which provide grey-scale images. This work makes use of images captured in visible range with color (RGB) information. We employ Gray-Level Co-Occurrence textural features and SVM classifiers for the task of fake iris detection. The best features are selected with the Sequential Forward Floating Selection (SFFS) algorithm. To the best of our knowledge, this is the first work evaluating spoofing attack using color iris images in visible range. Our results demonstrate that the use of features from the three color channels clearly outperform the accuracy obtained from the luminance (gray scale) image. Also, the R channel is found to be the best individual channel. Lastly, we analyze the effect of extracting features from selected (eye or periocular) regions only. The best performance is obtained when GLCM features are extracted from the whole image, highlighting that both the iris and the surrounding periocular region are relevant for fake iris detection. An added advantage is that no accurate iris segmentation is needed. This work is relevant due to the increasing prevalence of more relaxed scenarios where iris acquisition using NIR light is unfeasible (e.g. distant acquisition or mobile devices), which are putting high pressure in the development of algorithms capable of working with visible light.**

I. Introduction

Biometric systems have several advantages over traditional security methods based on something that you know (password, PIN) or something that you have (card, key, etc.) [1]. Users do not need to remember passwords or PINs (which can be forgotten) or to carry cards or keys (which can be stolen). Unfortunately, biometric systems are vulnerable to potential security breaches. Among them, *direct* or *spoofing* attacks are receiving particular attention [2]. In these attacks, the intruder tries to get access by using synthetically produced samples (e.g. gummy fingers or printed iris or faces) or by mimicking the behaviour of a genuine user (e.g. imitating the signature). An important issue in this type of attack is that no specific knowledge about the system internals is needed. It is carried out following the regular interaction mechanism with the sensor, and it occurs outside the digital limits of the system, so digital protection mechanisms (like encryption or watermarking) are not effective. Specific countermeasures against direct attacks have been proposed, which are usually classified into: i) *hardware-based*, in which an specific device for liveness detection is added to the sensor, and ii) *software-based*, in which fake samples are detected after the sample has been acquired with an standard sensor. Each solution has its own advantages and disadvantages. In general, hardware-based approaches have a higher fake detection rate. In this work, we will focus on software-based approaches (Figure 1), which have the advantages of being less expensive (they do not need extra hardware) and transparent to the user [2].

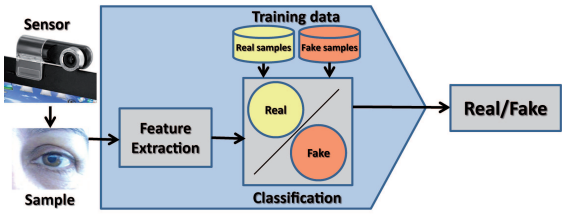

Fig. 1. System structure of a software-based liveness detection system.

Iris has been traditionally regarded as one of the most reliable and accurate biometric modalities available [3]. This has led to pay special attention to its vulnerabilities. The most common and simple spoofing attack is to present a high quality iris printed image. In our seminal work in 2008 [4], we captured a database of 800 fake images from 50 individuals and its corresponding real samples (ATVS-Flr DB), generated by presenting printed images of the originals to a commercial sensor[1]. We found in our experiments that about 40% of the fake images were enrolled by the system, and of those, more than 50% were matched successfully with its corresponding real sample. Also, as color contact lenses are becoming popular, another potential mean of spoofing is by using contact lenses with artificial textures printed onto them. Wei *et al.* [5] presented a database of 640 fake images from people wearing contact lenses from two manufacturers with 20 different types of artificial textures. They also proposed three measures (edge sharpness, Iris-Texton and three features from the co-occurrence matrix) to detect the printed contact lenses, reaching classification rates in the range of 76% to 100% (depending on the measure and manufacturer used). Real samples from the same people were not collected in these experiments; instead, they used real images from two open iris databases. In previous studies, Daugman proposed the use of spurious energy in the Fourier spectrum to detect printed iris patterns [6]. Lee *et al.* suggested the Purkinje image to detect fake iris [7], while He *et al.* used four image features (mean, variance, contrast and angular second moment) for this purpose [8]. There has been also research concerned with the synthesis of artificial images [9], accompanied by the release of datasets of synthetic iris images, such as the WVU-Synthetic Iris DB[2]. In 2013, LivDet-Iris 2013, the first Liveness Detection Iris Competition[3], was organized, with

[1] Available at http://atvs.ii.uam.es
[2] Available at www.citer.wvu.edu
[3] http://people.clarkson.edu/projects/biosal/iris/index.php

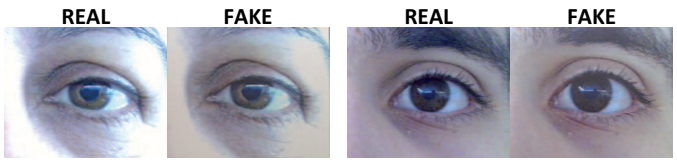

Fig. 2. Examples of real and fake images from the MobBIOfake database.

three academic participants. The datasets utilized (which are available to the public) included iris data from people wearing contact lenses and printouts of real iris images (with more than 4000 real and 3000 fake images). Classification rates averaged over the different types of data were in the range of 12%-25%. One difficulty of LivDet-Iris 2013 was the use of different contact lens manufacturers and different printers in the training and test data. Lastly, Galbally *et al*. proposed a general technique based on image quality features which allows detection of fake samples in image-based biometric modalities [2]. The latter followed a previous framework that we initiated with the use of trait-specific quality properties for liveness detection, including fingerprints [10], [11] and iris [12]. For the case of iris samples, the experiments reported in [2] achieved a classification rate of over 97% using the ATVS-Flr DB, and nearly 90% using synthetic iris images from the WVU-Synthetic Iris DB.

All the above-mentioned works have concentrated their efforts in data acquired with near-infrared (NIR) sensors, which provide gray-scale images. While this is the preferred choice of current commercial iris systems, visible wavelength imaging with color information are more appropriate for newer applications based on distant acquisition and 'on the move' capabilities, such as those using mobile devices [13]. In this direction, MobILive 2014, the 1st Mobile Iris Liveness Detection Competition [14] (currently underway) has provided the research community with a dataset [15] of 800 fake iris images (and its corresponding real images) acquired with a color webcam of a Tablet PC working in visible range (Figure 2). In this work, we evaluate the use of Gray-Level Co-Occurrence textural features [16], [17], [18] for the task of fake iris detection using such database. To the best of our knowledge, this is the first work evaluating spoofing attack using color iris images in visible range. We look for the best features by Sequential Forward Floating Selection (SFFS) [19], using SVM as classifier [20]. We demonstrate that the classification accuracy obtained from the luminance (gray scale) image can be considerably outperformed by an appropriate selection of features from the RGB channels, achieving a correct classification rate of over 96%. We also evaluate the extraction of GLCM features from the whole image vs. the extraction from selected (eye or periocular) regions only. The best classification rate is obtained when features are extracted from the whole image, highlighting that i) no accurate iris segmentation is needed, and ii) both the eye region and the surrounding periocular (skin) region provide valuable information for the task of fake image detection.

II. GLCM Textural Features

We employ the Gray Level Co-occurrence Matrix (GLCM) [16], [17], [18] for fake iris detection. The GLCM is a joint probability distribution function of gray level pairs in a given image $I(p,q)$. Each element $C(i,j)$ in the GLCM specifies the probability that a pixel with intensity value i occurs in the image $I(p,q)$ at an offset $d=(\Delta p, \Delta q)$ of a pixel with intensity value j. Usually the computation is done between neighboring pixels (i.e. $\Delta p = 1$ or $\Delta q = 1$). To achieve rotational invariance, the GLCM is computed using a set of offsets uniformly covering the 0-180 degrees range (e.g. 0, 45, 90 and 135 degrees). Once the GLCM is computed, various texture features are extracted and averaged across the different orientations. Let P_{ij} be the (i,j) entry in the GLCM. The features extracted are as follows:

Contrast: $f_1 = \sum_{i,j=0}^{N-1} P_{ij}\,(i-j)^2$

Dissimilarity: $f_2 = \sum_{i,j=0}^{N-1} P_{ij}\,|i-j|$

Homogeneity: $f_3 = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+|i-j|}$

Inverse Difference Moment: $f_4 = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$

Energy: $f_5 = \sum_{i,j=0}^{N-1} P_{ij}^2$

Maximum Probability: $f_6 = \max_{i,j} P_{ij}$

Entropy: $f_7 = \sum_{i,j=0}^{N-1} P_{ij}\,(-\ln P_{ij})$

GLCM mean: $f_8 = \mu_i = \sum_{i,j=0}^{N-1} iP_{ij}$

GLCM std: $f_9 = \sigma_i = \sqrt{\sum_{i,j=0}^{N-1} P_{ij}\,(i-\mu_i)^2}$

GLCM Autocorrelation: $f_{10} = \sum_{i,j=0}^{N-1} ijP_{ij}$

GLCM correlation: $f_{11} = \sum_{i,j=0}^{N-1} P_{ij}\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i\sigma_j}$

Cluster shade: $f_{12} = \sum_{i,j=0}^{N-1} P_{ij}\,((i-\mu_i)+(j-\mu_j))^3$

Cluster prominence: $f_{13} = \sum_{i,j=0}^{N-1} P_{ij}\,((i-\mu_i)+(j-\mu_j))^4$

In computing f_{11}, f_{12} and f_{13}, it must be considered that $\mu_i = \mu_j$ and $\sigma_i = \sigma_j$, due to the symmetry property of the GLCM [16]. Features f_1 to f_4 are related to contrast of the image, using weights related to the distance to the GLCM diagonal. Values of the diagonal show no contrast (pixel pairs with equal gray level), with contrast increasing away from the diagonal. Features f_5 to f_7 measure the regularity or order of the pixels in the image. Weights here are constructed based on how many times a pixel pair occur (given by P_{ij}). Lastly, features f_8 to f_{13} consist of statistics derived from the GLCM. All the extracted features are grouped into a single vector, which is used to model the image. We then use a SVM as classifier [20]. We have tested linear and polynomial (up to third-order) SVM kernels, with higher order kernels not considered due to the longer time required for computation.

III. Database and Protocol

For our experiments, we use the MobBIOfake database [15], which has been acquired in the framework of MobILive 2014, the 1st Mobile Iris Liveness Detection Competition [14]. MobILive 2014 is part of IJCB 2014, the International Joint

Conference on Biometrics[4]. MobBIOfake is derived from the iris images of the MobBIO database [21]. It is composed by 800 iris images from 100 volunteers and its corresponding fake copies, with a total of 1600 iris images. Samples were acquired with an Asus Eee Pad Transformer TE300T Tablet. The size of the color (RGB) iris images is of 200×240 pixels (height×width). Each volunteer contributed with 4 images of the two eyes. The fake samples were obtained from printed images of the original ones, captured with the same handheld device and in similar conditions. Here, we use the training dataset of MobBIOfake, released to the participants of MobILive 2014 to tune their algorithms and consisting of 400 iris images and its corresponding fake copies.

The task of fake biometric detection can be modeled as a two-class classification problem. The metrics used to evaluate the classification accuracy are: *False Acceptance Rate* (FAR), which accounts for the percentage of fake samples classified as real, and *False Rejection Rate* (FRR), which gives the percentage of real samples classified as fake. The average classification error (*Half Total Error Rate*) is then computed as HTER=(FAR+FRR)/2. Classification accuracy has been measured by cross-validation [22]. The database is divided into three disjoint sets, each set comprising one third of the available real images and their corresponding fake images. Two sets are chosen for training the classifier and one for testing, repeating the selection three times to consider the different possibilities. This yields to three classification errors, which are then averaged. We also evaluate different combinations of GLCM features for classification using SVMs. The best combination is found by Sequential Forward Floating Selection (SFFS) [19]. Given n features to combine, we employ as criterion value of the SFFS algorithm the HTER of the corresponding classifier trained with the n features.

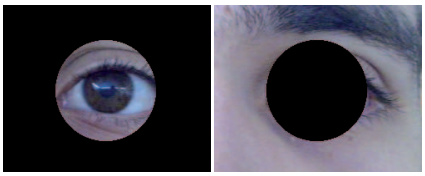

Fig. 3. Mask on (left) surrounding periocular and (right) eye region.

We also conduct detection experiments to localize the eye center position, which is used as input of the GLCM feature extraction algorithm, so as to extract GLCM features in the relevant eye/periocular region only. For this purpose, we employ our eye detection algorithm based on symmetry filters [23]. A circular mask of radius R=60 pixels is placed in the eye center, masking the corresponding outer (periocular) or inner (eye) region, depending on the experiment at hand (see Figure 3). The radius has been chosen empirically, based on the maximum radius of the outer (sclera) iris circle obtained by ground-truth annotation of the MobBIO database [23].

IV. Experiments and Results

Figure 4 shows the distribution of GLCM features on the real and fake iris images of the database (averaged between all images of each set, and normalized to the [0,1] range). The χ^2 distance between the real and fake histograms of each plot is also given. We evaluate three different cases in our experiments: i) GLCM features are extracted from the whole image (first row), ii) GLCM features are extracted from the eye region only (second row), and iii) GLCM features are extracted from the surrounding periocular region only (third row). See Figure 3 for further details of the different regions considered. GLCM features are extracted separately from the R, G and B color channels of the image, as well as from the corresponding (converted) gray scale image (named BW channel in our experiments). It can be observed in Figure 4 that there is difference between the histograms of real and fake images, justifying the applicability of GLCM features for the task of fake iris detection. Classification results of different combinations of GLCM features as selected with SFFS are given in Figure 5. We report the HTER values with linear and polynomial (up to third-order) SVM kernels. SFFS experiments are run separately on the R, G, B and BW channels of the image, with 13 available features in each case (columns 1-4). We also run SFFS by pooling together the R, G, and B features, having 13 × 3 = 39 features available for selection (column 5).

From Figure 5, we observe that a substantial performance improvement can be obtained in most cases with an appropriate combination of features, with the best SVM expert consistently being the linear one. The best individual channel (columns 1-4) is always the R channel. This is mirrored in Figure 4, where the biggest distances are always obtained between histograms of the R channel. On the other hand, the best classification accuracy is obtained when feature selection is done on the three RGB channels together (column 5), meaning that the three color channels contribute to the success of fake iris detection. It is relevant also that the luminance (BW channel) by itself is not able to provide good classification results (its performance is usually better than the G channel and similar to the B channel, but worse than the R channel). This points out the importance of using color information when available.

Considering the three different regions defined for feature extraction, the best classification accuracy is obtained when GLCM features from the whole image are used (first row of Figure 5). The classification error (HTER) in this case can go down to below 4% using a linear SVM (red curve of top right plot). For the other two cases of analysis, the lowest classification error with appropriate feature selection is in the range of 6-7%. Referring to the best configuration obtained (RGB channels, top left plot), when channels are properly chosen, the best performance can be obtained by combining a small number of features. After an initial sharp improvement in accuracy, it stabilizes at around 12 features. It is worth highlighting a significant range in the x-axis (12 to 35 features) where the performance is more or less constant. We give in Table I the features chosen by SFFS at three operating points of this range (12, 19 and 26 features selected). With 12 features, it is worth noting that they are equally chosen from the three color channels (4 from each one). For a higher number of selected features, SFFS has tendency towards choosing more features from the R channel (the best individual one), but in any case, the performance is not improved by selecting more than 12 features for classification.

Considering that the best performance is obtained with the

[4]http://ijcb2014.org

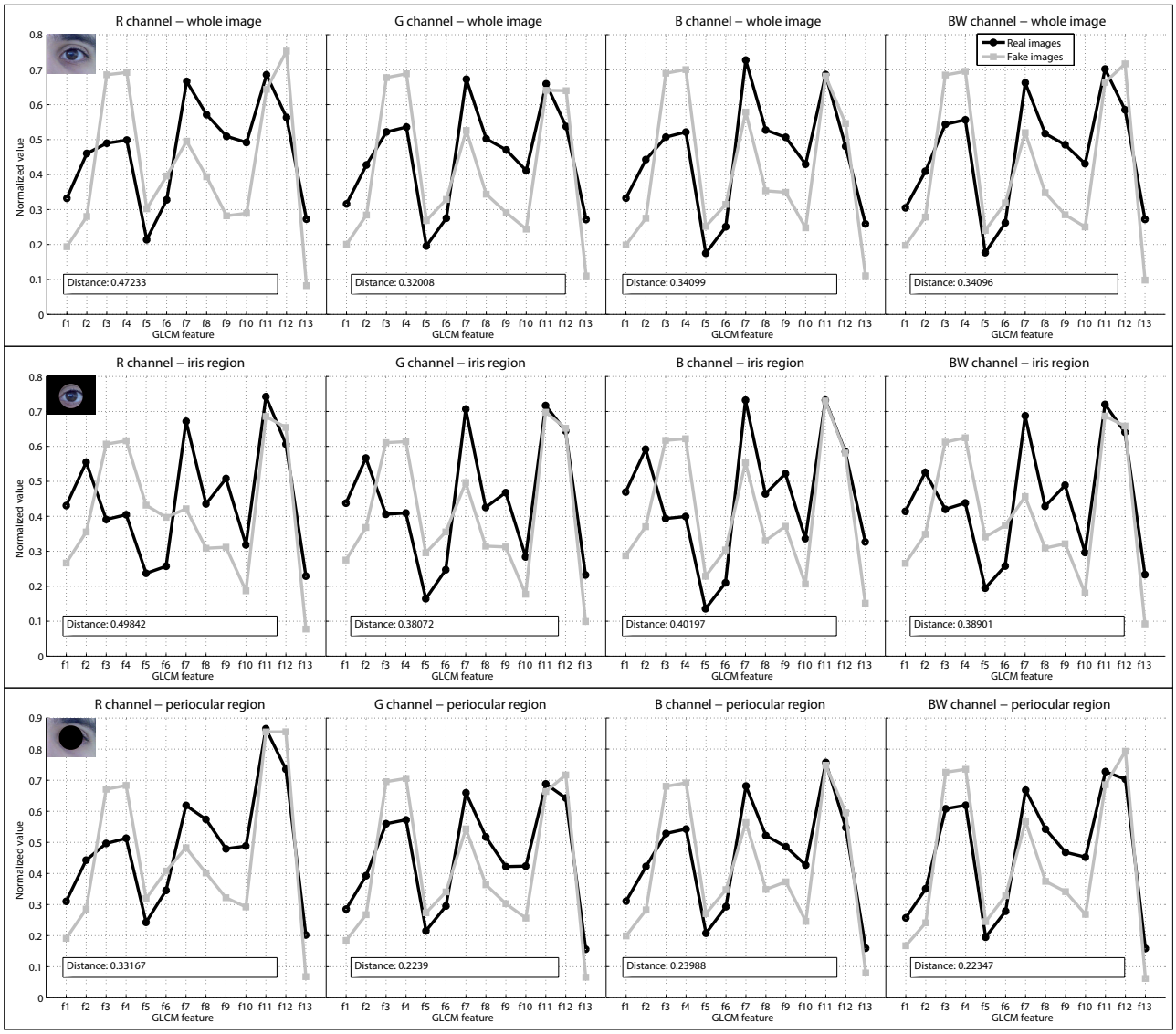

Fig. 4. Distribution of GLCM features for real and fake images (averaged over all available images of each class). The χ^2 distance between the histograms of real and fake images is also given. Top row: GLCM extracted from the whole image. Medium row: GLCM extracted from the iris texture region. Bottom row: GLCM extracted from the surrounding (periocular) region. For further details of the different regions considered, see Figure 3.

linear kernel (red line in Figure 5), we report in Figure 6 the three performance metrics (FAR, FRR and HTER) using this SVM only. Results of Figure 6 evidence that the FRR error (green curves) is usually higher than the FAR error (red curves), meaning that the features used have a higher tendency towards making errors when classifying a real image (FRR) rather than a fake image (FAR). While the system is doing its task of detecting fake images, it comes at the expense of increasing the number of users whose real samples are rejected. This has implications in terms of needing an operator who handles these exceptions (something which can be unfeasible in many applications) or increasing the annoyance of genuine users whose samples are rejected by the system. Both situations can have negative effects in the deployment of biometric systems [24]. This tendency, however, is compensated when selection is done on features from the three RGB channels (fifth column of Figure 6). Here, the difference between FRR and FAR is not so evident. This is another benefit with is added to the fact observed above that the best performance is obtained by combining features from the three RGB channels.

V. Conclusion

Previous research on fake iris detection has concentrated their efforts in data acquired with near-infrared sensors (providing gray-scale images), which is the preferred choice of current commercial imaging systems [13]. Here, we evaluate the use of GLCM textural features [16], [17], [18] and SVM classifiers [20] for the task of fake iris detection with color images in visible range, in which, to the best of our knowledge, is the first study using data acquired in such conditions. The proposed method is tested on a database of 800 fake and its corresponding real images acquired with a webcam of a Tablet PC [14]. The best features for fake detection are selected with SFFS [19]. The classification accuracy by using the luminance (gray scale) channel is substantially outperformed by selecting

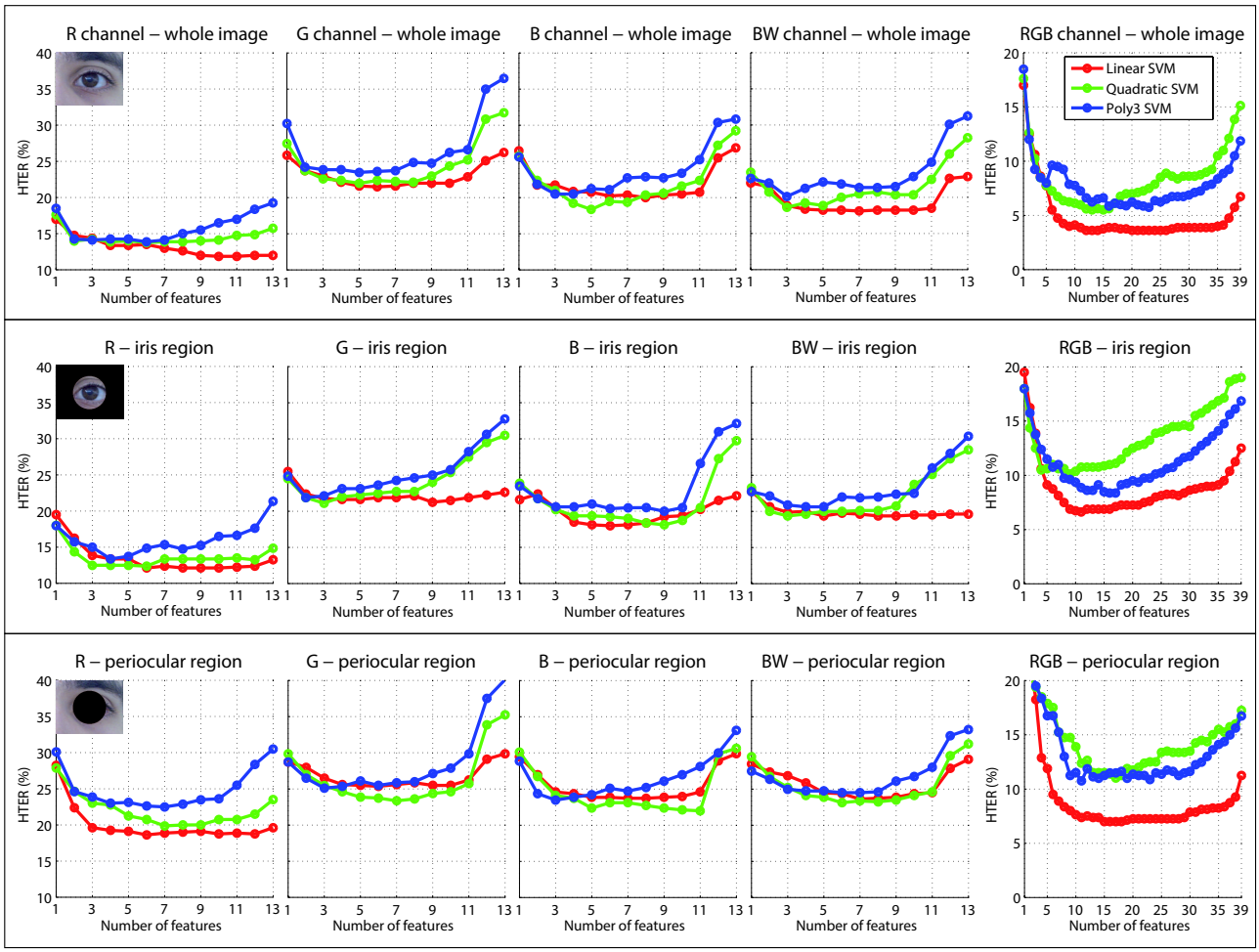

Fig. 5. Classification results (HTER) for an increasing number of textural features (selected with SFFS) with different SVM kernels. Top row: GLCM extracted from the whole image. Medium row: GLCM extracted from the iris texture region. Bottom row: GLCM extracted from the surrounding (periocular) region. For further details of the different regions considered, see Figure 3.

	R channel													G channel												B channel														
# feats	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	**HTER**	**FRR**	**FAR**
12							x		x	x			x				x		x		x												x	x	x				3.62	4.24	3.01
19		x		x				x		x	x		x			x	x			x	x	x	x	x				x	x				x	x		x			3.75	3.49	4.01
26	x	x	x	x	x			x		x	x	x	x		x	x	x	x	x			x	x				x	x	x		x				x	x		x			3.63	3.24	4.01

TABLE I. FEATURES SELECTED BY SFFS FOR DIFFERENT SIZES OF THE TARGET SET (LINEAR SVM). REPORTED RESULTS CORRESPOND TO THREE OPERATING POINTS OF THE TOP RIGHT PLOT OF FIGURE 6.

features from the RGB channels, reaching a correct classification rate of over 96%. In addition, the R channel is shown to be the best individual channel, outperforming the G and B channels, as well as the luminance channel. We also analyze the extraction of GLCM features from the whole image vs. the extraction from selected (eye or periocular) regions only. The best case performance is obtained when features are extracted from the whole image, pointing out that both the eye and the surrounding periocular region contribute to the success of the fake detection task. An added advantage is that no accurate iris segmentation is needed.

Currently, we are extending this same study to NIR images. Since the skin surrounding the eye is usually over-illuminated with NIR illumination, we speculate that there might be differences in the results with this type of images. We are also working on including other type of fake data in our analysis, such as contact lenses [5].

ACKNOWLEDGMENTS

F. A.-F. thanks the Swedish Research Council and the EU for funding his postdoctoral work. Authors acknowledge the CAISR program of the Swedish Knowledge Foundation, the EU BBfor2 project and the EU COST Action IC1106.

REFERENCES

[1] A. K. Jain and A. Kumar, *Second Generation Biometrics*, chapter Biometrics of Next Generation: An Overview, Springer, 2010.

[2] J. Galbally, S. Marcel, J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE TIP*, vol. 23, no. 2, pp. 710–724, Feb 2014.

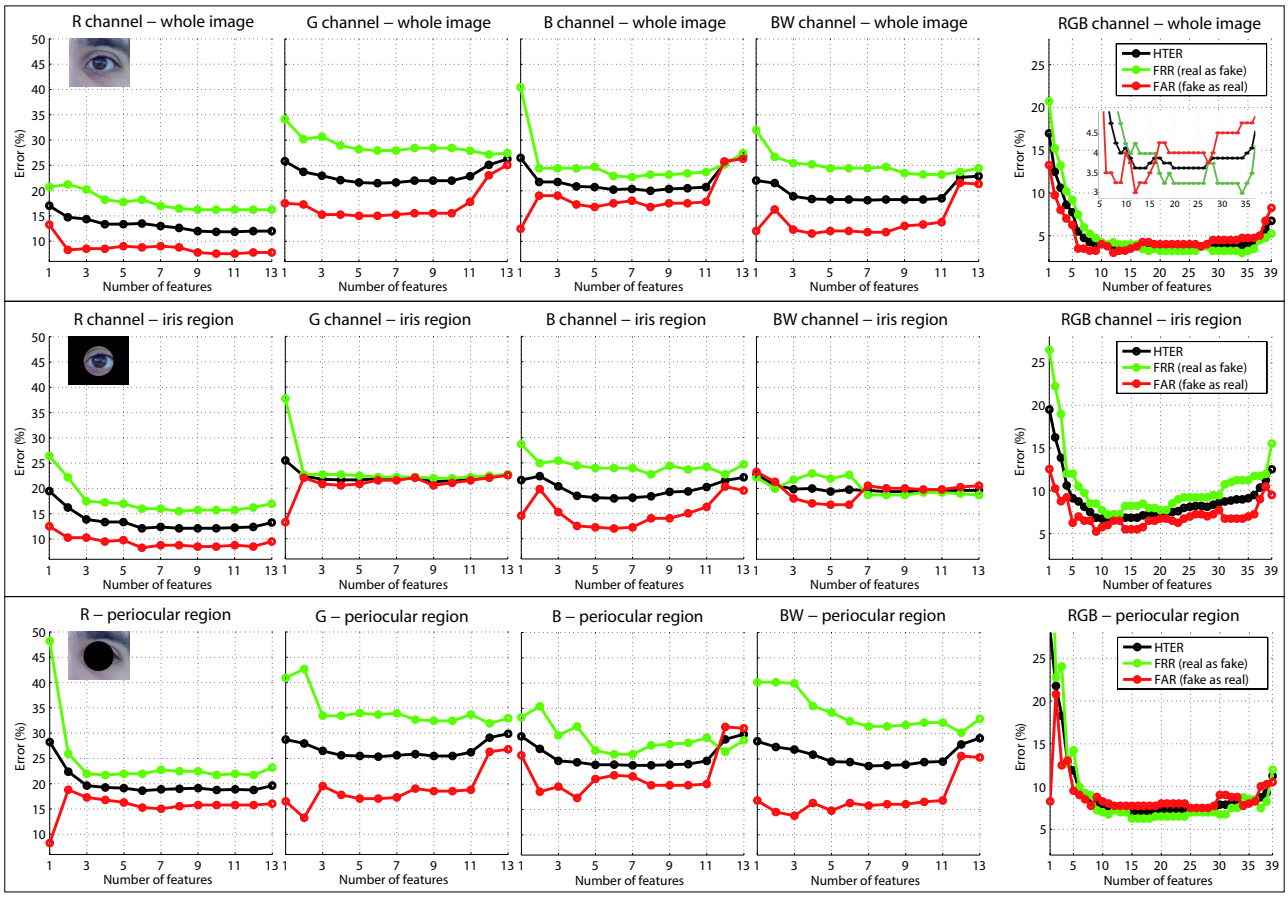

Fig. 6. Detail of classification results (FAR, FRR and HTER) for an increasing number of textural features (selected with SFFS) using a linear SVM kernel. Top row: GLCM extracted from the whole image. Medium row: GLCM extracted from the iris texture region. Bottom row: GLCM extracted from the surrounding (periocular) region. For further details of the different regions considered, see Figure 3.

[3] K.W. Bowyer, K. Hollingsworth, P.J. Flynn, "Image understanding for iris biometrics: a survey," *CVIU*, vol. 110, pp. 281–307, 2007.

[4] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, J. Ortega-Garcia, "Direct attacks using fake images in iris verification," *Proc. BIOID*, Springer LNCS-5372, pp. 181–190, 2008.

[5] Z. Wei, X. Qiu, Z. Sun, T. Tan, "Counterfeit iris detection based on texture analysis," *Proc. ICPR*, pp. 1–4, 2008.

[6] John Daugman, "Demodulation by complex-valued wavelets for stochastic pattern recognition," *Intl J Wavelets, Multi-resolution and Information Processing*, vol. 1, pp. 1–17, 2003.

[7] E. C. Lee, K. R. Park, J. Kim, "Fake iris detection by using purkinje image," *Proc. ICB*, Springer LNCS-3832, pp. 397–403, 2006.

[8] X. He, S. An, P. Shi, "Statistical texture analysis-based approach for fake iris detection using support vector machines," *Proc. ICB*, Springer LNCS-4642, pp. 540–546, 2007.

[9] S. Shah and A. Ross, "Generating synthetic irises by feature agglomeration," in *Proc. IEEE ICIP*, Oct 2006, pp. 317–320.

[10] J. Galbally, F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Elsevier FGCS*, vol. 28, no. 1, pp. 311–321, 2012.

[11] J. Galbally, F. Alonso-Fernandez, J. Fierrez-Aguilar, J. Ortega-Garcia, "Fingerprint Liveness Detection Based on Quality Measures," *Proc. IEEE BIDS*, 2009.

[12] J. Galbally, J. Ortiz-Lopez, J. Fierrez, J. Ortega-Garcia, "Iris liveness detection based on quality related features," in *Proc. ICB*, pp. 271–276, 2012.

[13] K. W. Bowyer, K. P. Hollingsworth, P. J. Flynn, "A survey of iris biometrics research: 2008-2010," in *Handbook of Iris Recognition*, M. J. Burge and K. W. Bowyer, Eds., Advances in Computer Vision and Pattern Recognition, pp. 15–54. Springer London, 2013.

[14] MobILive 2014, "The 1st mobile iris liveness detection competition - http://mobilive2014.inescporto.pt," 2014.

[15] A. F. Sequeira, J. Murari, S. Cardoso, "Iris liveness detection methods in mobile applications," *Proc VISAPP*, vol. 3, pp. 133–139, 2014.

[16] R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE TSMC*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973.

[17] L.-K. Soh and C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices," *IEEE TGRS*, vol. 37, no. 2, pp. 780–795, Mar 1999.

[18] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can J Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[19] P. Pudil, J. Novovicova, J. Kittler, "Flotating search methods in feature selection," *Patt Recogn Letters*, vol. 15, pp. 1119–1125, 1994.

[20] Vladimir N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[21] A. F. Sequeira, J. C. Monteiro, A. Rebelo, H. P. Oliveira, "Mobbio: a multimodal database captured with a portable handheld device," *Proc VISAPP*, vol. 3, pp. 133–139, 2014.

[22] R. Duda, P. Hart, D. Stork, *Pattern Classification - 2nd Edition*, 2004.

[23] F. Alonso-Fernandez and J. Bigun, "Eye detection by complex filtering for periocular recognition," *Proc. IWBF*, 2014.

[24] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Security and Privacy*, vol. 10, no. 6, pp. 52–62, 2012.