An improved simple feature set for face presentation attack detection

Anna Denisova[1,2]

[1]Samara National Research University
Moskovskoye shosse 34,
443086, Samara, Russia
[2]Image Processing Systems Institute – Branch of the Federal Scientific Research Centre
"Crystallography and Photonics" of
Russian Academy of Sciences
Molodogvardeiskaya st. 151
443001, Samara, Russia

denisova_ay@geosamara.ru

ABSTRACT

Presentation attacks are weak points of facial biometrical authentication systems. Although several presentation attack detection methods were developed, the best of them require a sufficient amount of training data and rely on computationally intensive deep learning based features. Thus, most of them have difficulties with adaptation to new types of presentation attacks or new cameras. In this paper, we introduce a method for face presentation attack detection with low requirements for training data and high efficiency for a wide range of spoofing attacks. The method includes feature extraction and binary classification stages. We use a combination of simple statistical and texture features and describe the experimental results of feature adjustment and selection. We validate the proposed method using WMCA dataset. The experiments showed that the proposed features decrease the average classification error in comparison with the RDWT-Haralick-SVM method and demonstrate the best performance among non-CNN-based methods.

Keywords

Presentation attack detection, feature extraction, depth map, thermal data, infrared data, WMCA, SVM, RDWT-Haralick-SVM, MC-CNN.

1. INTRODUCTION

Recent research in face recognition systems vulnerability revealed high Impostor Attack Presentation Match Rates (IAPMR) for almost all kinds of face recognition systems [Bha19]. This fact demonstrates the crucial need of the development of presentation attack detection (PAD) mechanisms to protect face recognition systems. The general scenario of presentation attack is to demonstrate a fake image of the bona fide person to the authentication system. Presentation attack instruments may include printed photos, masks or screen photos that can be easily reproduced due to high accessibility of the target person's photos in social networks and other open data sources.

To prevent face recognition systems from presentation attacks five basic approaches were developed. The motion based approach exploits an additional analysis of face movements in the video sequence to extract liveness information [Anj11; Fre12; Liu09]. For example, the eye blinking in the video distinguishes the real person's image from printed photos. The motion based PAD methods suffer from long time of data registration and require in most cases additional actions from the user. The second approach is texture based methods [Pen18; Pen20; Du21]. Texture based methods suppose that the real face image has significantly different textural properties from the fake image reproduced using printing or screen devices. However, this kind of method may not be effective in the case of more complex attacks such as 3D masking. The image quality based methods belong to the third approach that calculates different quality measures for fake and real images [Gal14; Wen15]. Usually low cost spoofing devices lead to image quality degradation that is monitored by this group of methods. Nevertheless, high quality spoofing cannot be detected by these methods. The fourth approach is

multimodal spoofing detection. The main modality used for PAD is RGB color images. The other modalities such as infrared images, depth maps or thermal images provide useful additional information for liveness detection and can improve the PAD detection for many kinds of sophisticated attacks as well as for the novel PAD instruments [Wan20; Abd21; Kow20]. For example, 3D-latex and silicone masks can be easily detected using thermal data due to the difference in thermal characteristics of the artificial materials in comparison to the real faces [Kow20]. The last approach, that is worth mentioning, is deep learning. Deep learning algorithms are based on artificial neural networks and use tones of images for training. They can be used in both multimodal and single modal cases. At the current time, deep learning methods provide the best quality of spoofing detection [Wan19; Geo19; Bre19]. However, the need for large training datasets makes these methods inconvenient for launching on new devices when it is impossible to collect a lot of training data.

In this article, we propose a multimodal PAD method based on simple handcrafted features that can be easily calculated for new imaging devices. The features obtained for each particular sensor are concatenated and then processed together by the classifier. The method handles the proposed features using one of the classical binary classification methods such as linear support vector machines classifier (linear SVM) and random forest classifier (RF). These classification methods allow training with small training sets. Thereby the method combines the universality of feature extraction for new modalities and the simplicity of training provided by classical binary classification approaches.

We compared our method with two algorithms such as RDWT-Haralick-SVM and MC-CNN which have shown the best spoofing detection performance according to [Geo19]. The RDWT-Haralick-SVM algorithm leans on linear SVM classifier and texture features. MC-CNN is a deep learning algorithm that uses an artificial neural network for feature extraction and classification. Our method outperformed the RDWT-Haralick-SVM algorithm and gave closer results to MC-CNN. Although MC-CNN demonstrates higher performance, it uses an extremely large dataset for training whereas the proposed method can be trained only with hundreds of images. Thereby, the proposed algorithm is more suitable in cases when training images are hard to obtain.

The rest of the paper is organized as follows. Section 2 describes the proposed feature set. Section 3 provides the experimental results of feature set adjustment and comparison of the proposed method with the baseline algorithms such as RDWT-Haralick-SVM and MC-CNN.

2. PROPOSED METHOD

2.1 Feature Extraction

There are several input images $X, Y_1, Y_2, ..., Y_N$ of the target face. The image X is a grayscale image obtained by RGB camera and $Y_1, Y_2, ..., Y_N$ are the images obtained by different additional sensors (sensor images), for example, depth map or infrared image. All images are geometrically consistent and have the same size $I \times J$ pixels. Input images are linearly contrasted in the range $[0, 255]$. The features are calculated in two stages. Firstly, we calculate feature vector $f_n \in R^M$ for a pair of grayscale and sensor image X, Y_n. Finally, all feature vectors are concatenated $\left(f_1^T, f_2^T, ..., f_N^T\right)^T \in R^{NM}$.

There are two key ideas behind the feature extraction process. First, if there is no attack the images X and Y_n have to demonstrate the same face contours. Second, if there is an attack the images X and Y_n have to significantly differ from each other in both statistical and textural aspects. Using these ideas we constructed our feature set from 6 basic feature groups which were examined in our paper [Den21]. Even with these feature groups we obtained better spoofing detection. However, texture features are a very powerful instrument of image analysis and, in this paper, we add Haralick features as the seventh basic feature group. The basic feature groups are presented in Table 1. Following by our intuitions we consider some additional feature groups as well. These features are optional and may be included or omitted depending on experimental results. The list of additional feature groups is provided in Table 2.

The first basic feature group (Var) includes the variance of the sensor image. To compute the second basic feature group (AvgArr) we divide the sensor image on $K \times K$ regions and calculate the average brightness for each region.

The third basic group (GradYH) is a histogram bins $H_{Yn}(q), q = 1, ..., Q$ of the gradient amplitude G_{Yn} for sensor image. We take into account only values of the gradient amplitude from 0 to 30 and the number of bins is supposed to be $Q = 20$. This group of feature group allows us to distinguish between print and replay attacks when the sensor images do not have contours of the face and, therefore, demonstrate low gradient amplitude values.

The fourth basic group (CorrY) includes the coefficients of sensor image correlation. If $B_n(p_1, p_2)$ is an autocorrelation function of sensor image Y_n, the values of $B_n(p_1, p_2)$ for particular

p_1, p_2 are the features values. In this paper, we consider correlation in four directions $B_n(-1,0)$, $B_n(1,0)$, $B_n(0,1)$ and $B_n(0,-1)$. Usually, the correlation for fake images is higher than for bona fide images.

Notation	Images needed for computation	Number of features in group
Var	Y_n	1
AvgArr	Y_n	$K^2 = 9$
GradYH	Y_n	$Q = 20$
CorrY	Y_n	4
GradXH	X, Y_n	$Q = 20$
CorrXY	X, Y_n	1
H13	Y_n	208

Table 1. Basic feature groups

Notation	Images needed for computation	Number of features in group
StatY	Y_n	4
HistY	Y_n	$Q = 20$
GradDirYH	Y_n	$Q = 20$
GradDirXH	X, Y_n	$Q = 20$

Table 2. Additional feature groups

The fifth basic feature group (GradXH) is a histogram $H_{X(n)}(q), q = 1,...,Q$ of the gradient amplitude G_X of the grayscale image X in contour points of the sensor image Y_n. The contour points $i, j \in \Omega_{Y(n)}$ are the points detected by Canny detector in the image Y_n. The histogram $H_{X(n)}(q)$ is computed only for values $G_X(i,j), (i,j) \in \Omega_{Y(n)}$. We compute the histogram $H_{X(n)}(q)$ only for the range from 0 to 128 because these values of the gradient amplitude are more informative and correspond to local brightness variations within the face area. The number of bins is 20. These features aim to evaluate the correspondence between the facial contours on the sensor and grayscale images.

The sixth basic feature group (CorrXY) is the mutual correlation of the gradients G_X and $G_{Y(n)}$.

The seventh basic feature group (H13) is a concatenation of 13 Haralick features obtained for different parts of the sensor image. We divide the sensor image Y_n into $W \times W$ regions and compute 13 Haralick features for each region: Angular Second Moment, Correlation, Inverse Difference Moment, Sum Variance, Entropy, Difference Entropy, Contrast, Sum of Squares: Variance, Sum Average, Sum Entropy, Difference Variance, Info. Measure of Correlation 1 and Info. Measure of Correlation 2. The names of features correspond to the ones in the article [Har73]. In this paper we used *W=4*.

As for additional feature groups, they were designed intuitively to improve the classification performance. We added the features explaining the statistical properties of the sensor image and the gradient phase based features. The feature group StatY is the mean, the median, the range and the maximum value obtained for the sensor Y_n image. The feature group HistY is the histogram $H_{Y(n)}(q), q = 1,...,Q$ of the sensor image Y_n. For this histogram we use the full range of pixel values from 0 to 255 and $Q = 20$. The next feature group (GradDirYH) is a histogram $HDir_{Yn}(q), q = 1,...,Q$ of the gradient phase $GDir_{Yn}$ for the sensor image. The histogram range is from -180 degrees to 180 degrees. The number of bins remained the same $Q = 20$. The last feature group is the gradient phase histogram (GradDirXH) $HDir_{X(n)}(q), q = 1,...,Q$ for image X in contour points of the sensor image Y_n. This feature group is calculated in a similar way that is for GradXH feature group. The only difference is using the gradient phase instead of the amplitude to build the histogram. The range of the histogram bins varies from -180 degrees to 180 degrees. The number of bins is equal to 20.

The proposed features have several advantages. First of all, they are simple in computation. Secondly, they are enough effective even for images with degraded quality when the sensor image Y_n has a lower resolution than the grayscale image X.

2.2 Classification

We consider the presentation attack detection as a binary classification problem. The bona fide presentations are classified in class 1. The fake images are classified in class 0. To perform classification we use one of three algorithms: linear SVM (Support Vector Machines with linear kernel) [Chr00] and RF (Random Forest) [Kul12]. These algorithms are able to work with small training sets containing several hundreds of images.

To find the parameters of each algorithm we produced a grid search with optimization of the following standard measures: attack presentation classification error rate (APCER), bona fide presentation classification error rate (BPCER) and average classification error rate (ACER). The APCER corresponds to False Positive Rate. BPCER

is the same as False Negative Rate. ACER is the average of APCER and BPCER.

For linear SVM, we tested penalty parameter C in the range [0,1] with step 0.0001 and the cost matrix A. The cost matrix was chosen as one of three matrixes:

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix},$$

where the cost of bona fide class is 1, 2 and 4.

For the RF algorithm, we optimized the tree number N and cost matrix A. We tested N from 10 to 200 with the step 10.

The cost matrixes and algorithm parameters' ranges were selected during the preliminary experiments.

3. EXPERIMENTAL RESEARCH

3.1 Dataset

We evaluated our method using the Wide Multi-Channel Presentation Attack Database (WMCA) [Geo19]. WMCA includes 1679 videos which contain 347 bona fide presentations and 1332 attacks. The attacks are grouped in the following categories:

– "fake head" when the heated mannequin was used instead of real face;

– "print attack" when printed photo of the target person is demonstrated to the camera;

– "replay" when the screen with the person's portrait is demonstrated to the camera;

– "rigid mask" when the person wears the rigid mask made in a handcrafted manner from rigid materials;

– "flexible mask" when the person wears the soft silicon mask;

– "paper mask" when the person wears a handcrafted paper mask;

– "glasses" when the person wears paper or funny eyes glasses.

WMCA database incorporates data captured in four modalities: RGB color data, depth map, infrared data and thermal data. The first three modalities were captured using Intel RealSense SR300. Thermal data were collected using Seek Thermal Compact PRO. Each image in the dataset is presented in four channels color image intensity (C), depth (D), infrared image (I) and thermal data (T). The images are geometrically aligned and have 128 × 128 pixels in size. The examples of four image channels are presented in Figure 1.

In [Geo19] the authors used approximately 50 frames per video for data augmentation that is 83950 images in total. We will refer to this case as the full dataset. In our case, we do not need an augmentation to train classifiers that is why we used only one frame per video that is 1679 images in total. We will refer to this case as the basic dataset. Most of our experiments are provided for the basic dataset.

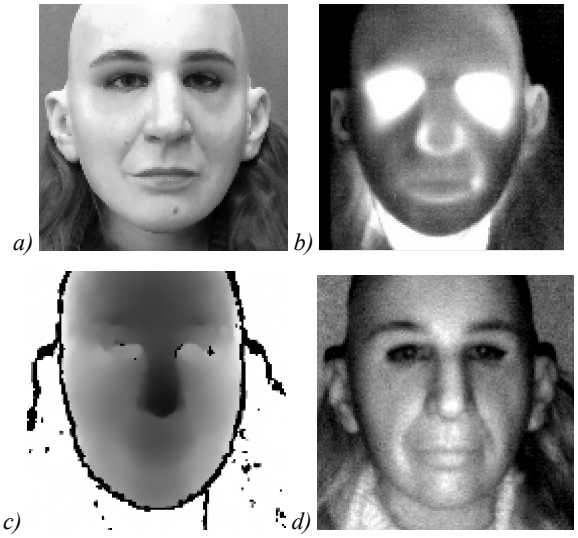

Figure 1. The example of data in CDIT channels for fake head attack: a) color image intensity, b) thermal data, c) depth, d) infrared

3.2 Baseline Methods

We focus on the two methods that demonstrated the best results in the paper [Geo19]. The first one is RDWT-Haralick-SVM method proposed in [Ewa20]. This method is based on RDWT-Haralick features and linear SVM classification. To compare our method with RDWT-Haralick-SVM in the same conditions we reimplemented this method. The final feature vector in our implementation is a concatenation of RDWT-Haralick features obtained for each image channel separately. The RDWT-Haralick features for one image channel are obtained in the same way as described in [Geo19].

The second baseline method is MC-CNN developed in [Geo19]. This method is based on LightCNN deep learning model proposed in [Wu18]. The MC-CNN extends pre-trained LightCNN for four channels and applies additional training to some layers of the model. A detailed description of the MC-CNN architecture is given in [Geo19].

3.3 Experiment protocol

In this paper, we follow the grandtest protocol described in [Geo19]. We divide datasets into training, test and validation (development) sets in the same manner. Thus, training, test and validation sets do not have common bona fide presentations and demonstrate the almost equal distribution of presentation attack instruments.

We provide two series of experiments. In the first series, we evaluate only the basic feature set and define the basic performance of the method in different conditions. In the second series, we test our

intuitions on further performance improvement using additional features and adjusting some basic features.

3.4 Channel Selection

In the first experiment, we assessed the optimal channel set. We tested the algorithm using basic dataset. The results are shown in Table 3. The best performance is demonstrated for all channels (CDIT) and corresponds to an ACER value equal to 2.91%. It is less than the ACER=3.44% obtained by RDWT-Haralick-SVM on the full dataset. Thus, the proposed basic feature set improves the classification results obtained for CDIT channels even for small training sets. Further, we provide the results only for CDIT channel configuration.

3.5 Classifier selection

To investigate classifiers' performance and compare the results with the baseline methods we carried out experiments with the basic and full datasets. The results for the basic dataset are shown in Table 4. The proposed method demonstrates the best results for linear SVM classifier (ACER=2.91%). The RDWT-Haralick-SVM achieves ACER bigger in 1.43 times. Therefore, the proposed method in combination with linear SVM improves spoofing detection for the small training sets in comparison to RDWT-Haralick-SVM.

The classification results for the full dataset are given in Table 5. The proposed algorithm achieves the best result with ACER=1.18% using linear SVM and it still outperforms RDWT-Haralick-SVM (ACER=3.44%) but loses to MC-CNN (ACER=0.3%). Nevertheless, we suppose that MC-CNN is impossible to train in the case of small datasets including hundreds of images whereas the proposed algorithm is successfully trained in this case.

3.6 Additional Feature Set Testing and Basic Feature Set Parameters Adjustment

The second series of experiments aimed to check the following our intuitions about feature improvement:

1) RGB data and gradient phase can improve the classification results. These two heuristic scenarios correspond to the cases when RGB channels are considered as additional independent modalities and when the gradient phase histograms are taken into account as additional features GradDirYH and GradDirXH (see Table 2).

Channels	Optimal hyperparameters		Validation set		Test set		
	C	A	APCER, %	BPCER, %	APCER, %	BPCER, %	ACER, %
CDIT	**0.0055**	**A_3**	**3.34**	**1.85**	**4.07**	**1.74**	**2.91**
CDT	0.0119	A_3	4.90	0	7.01	6.96	6.99
CTI	0.0111	A_3	4.45	0.93	5.20	1.74	3.47
CDI	0.0015	A_3	11.58	7.41	9.28	0	4.64
CT	0.0439	A_3	7.13	0	10.86	1.74	6.30
CD	0.0319	A_3	10.02	13.89	12.90	6.09	9.49
CI	0.0856	A_3	6.46	3.70	6.11	8.70	7.40

Table 3. Classification errors for the test set for different input channels (basic dataset, linear SVM)

Method	Optimal hyperparameters	APCER, %	BPCER, %	ACER, %
Basic feature set + SVM	**C = 0.0055, A_3**	**4.07**	**1.74**	**2.91**
Basic feature set + RF	N=80, A_2	5.88	7.83	6.85
RDWT-Haralick+SVM	C=0.0002, A_3	8.37	0	4.18

Table 4. Classification errors for the test set in the case of CDIT data and basic dataset of 1679 images

Method	Optimal hyperparameters	APCER, %	BPCER, %	ACER, %
Basic feature set + SVM	C = 1.1327, A_3	0.11	2.24	1.18
Basic feature set + RF	N=80, A_2	3.23	6.50	4.87
RDWT-Haralick+SVM [Geo19]	-	6.39	0.49	3.44
MC-CNN [Geo19**]**	**-**	**0.60**	**0.0**	**0.30**

Table 5. The classification errors for the test set in the case of CDIT data and full dataset of 83950 images

2) the optimal number of regions used for AvgArr feature calculation can be selected. The feature AvgArr is composed of statistics obtained for $K \times K$ regions of the image. However, the preliminary test of the algorithm included only results for $K = 3$ while the other values of K may provide better results.

3) face region masking can improve the results. The face region and the background probably contain different information about spoofing methods. Thus, it is possibly better to calculate features in these regions separately.

We checked these three hypotheses for basic and extended feature sets. The last one includes basic features (see Table 1) and two additional features HistY and StatY (see Table 2). The experiments were carried out using basic dataset (1679 images) and linear SVM classifier.

Table 6 shows the results of our first intuition. As we can see from Table 6, the RGB channels do not improve the classification results. On the contrary, the gradient phase histograms improved the average classification error by almost 1%. Besides the extended feature set demonstrated a little bit better result (ACER=2.04%) than the basic feature set (ACER=2.37).

Table 7 demonstrates the results for the second intuition. The experiments showed that the optimal value of K is 8. Moreover, the extended feature set demonstrated a significant decrease in ACER (1.69%) for optimal K.

Table 8 demonstrates the results for different ways of masking. The separation of an image on the face and non-face regions achieves the best classification error. Besides, the result obtained for the extended feature set (ACER=1.70%) is better than the one for the basic feature set (ACER=2.36%).

To summarize, we defined three strategies for feature improvement. The first strategy is the use of gradient phase histograms. The second strategy is the use of $K = 8$ for AvgArr computation. And the last one is to separate feature extraction for face and non-face regions by means of masking.

Features	Channels	Optimal hyperparameters	ACER%	APCER%	BPCER%
Basic	CDIT	C= 0.0055, A_3	2.91	4.07	1.74
Extended	CDIT	C= 0.0032, A_3	3.00	3.39	2.61
Basic	CDIT+RGB	C= 0.0060, A3	4.26	6.79	1.74
Extended	CDIT+RGB	C= 0.0026, A3	3.81	5.88	1.74
Basic + GradDirYH + GradDirXH	CDIT	C = 0.0218, A3	2.37	4.75	0
Extended + GradDirYH + GradDirXH	CDIT	C = 0.0019, A3	2.04	4.07	0

Table 6. Classification errors for the test set and for the first hypothesis (linear SVM, basic dataset)

Features	K	Channels	Optimal hyperparameters	ACER%	APCER%	BPCER%
Basic	2	CDIT	C=0.0058, A3	3.88	4.30	3.48
Basic	3	CDIT	C=0.0055, A_3	2.91	4.07	1.74
Basic	4	CDIT	C=0.0057, A3	3.66	3.85	3.48
Basic	**8**	**CDIT**	**C=0.0178, A3**	**2.24**	**3.62**	**0.87**
Extended	2	CDIT	C=0.0035, A3	3.00	3.39	2.61
Extended	3	CDIT	C=0.0032, A_3	3.00	3.39	2.61
Extended	4	CDIT	C=0.0035, A3	3.00	3.39	2.61
Extended	**8**	**CDIT**	**C=0.0036, A3**	**1.69**	**3.39**	**0**

Table 7. Classification errors for the test set and for the second hypothesis (linear SVM, basic dataset)

Features	Mask	Optimal hyperparameters	ACER%	APCER%	BPCER%
Extended	One elliptical face-mask	C=0.0046, A_3	2.92	4.98	0.87
Extended	**Two masks for face and non-face regions**	**C=0.0020, A3**	**1.70**	**3.39**	**0**
Extended	No mask	C=0.0032, A3	3.00	3.39	2.61
Basic	One elliptical face-mask	C=0.0035, A3	2.38	4.75	0
Basic	**Two masks for face and non-face regions**	**C=0.0050, A3**	**2.36**	**3.85**	**0.87**
Basic	No mask	C=0.0055, A3	2.91	4.07	1.74

Table 8. Classification errors for the test set and for third hypothesis (linear SVM, basic dataset)

Features	Gradient phase histograms	AvgArr, K	Optimal hyperparameters	ACER%	APCER%	BPCER%
Basic	+	**8**	**C=0.0121, A_3**	**2.26**	**4.52**	**0**
Extended	+	8	C=0.0054, A_3	1.81	3.62	0
Basic	-	3	C=0.0055, A_3	2.91	4.07	1.74
Extended	-	3	C=0.0032, A_3	3.00	3.39	2.61
Basic	+	3	C=0.0218, A3	2.37	4.75	0
Extended	+	3	C=0.0019, A3	2.04	4.07	0
Basic	-	8	C=0.0178, A3	2.24	3.62	0.87
Extended	**-**	**8**	**C=0.0036, A3**	**1.69**	**3.39**	**0**

Table 9. Classification errors for the test set in the case of first and second feature improvement strategies (CDIT channels, basic dataset of 1679 images)

Features	Gradient phase histograms	AvgArr, K	Masking	Optimal hyperparameters	ACER%	APCER%	BPCER%
Basic	+	8	Two masks for face and non-face regions	C=0.0025, A3	1.69	3.39	0
Extended	-	8	Two masks for face and non-face regions	C=0.0019, A3	1.58	3.17	0

Table 10. Classification errors for the test set in the case of best feature improvement strategies (CDIT channels, basic dataset of 1679 images)

Table 9 provides the results for the simultaneous use of first two strategies. For extended feature set, the best choice is to use the second strategy only. For basic dataset, the best choice is using both first and second strategies simultaneously.

The further evaluation of the best strategies is given in Table 10. In both cases, simultaneous use with the third strategy provides the decrease in classification error.

In the end, we may conclude that all of the explored intuitions lead us to classification performance improvement. The best strategies shown in Table 10 demonstrate a decrease of classification error almost 1.84 times in comparison to the basic feature set with the default parameters.

4. CONCLUSION

In this paper, we propose a method for multimodal face presentation attack detection. The method is based on simple statistical and texture feature extraction and classical binary classification methods. The main advantages of the proposed method are the universality in terms of sensors used to obtain data and the simplicity of training on small training sets. Therefore, the main use cases of the proposed method are when the training data are limited or hard to achieve and when the new sensor is introduced into the PAD system.

We consider the basic method implementation including the basic set of features and one of two classification algorithms linear SVM and RF and the additional feature set with some heuristic strategies of feature improvement. The proposed method was verified using the WMCA database. The comparison of basic method implementation with two baseline methods showed that the proposed method gives better results than the RDWT-Haralick-SVM method, however, it loses to MC-CNN method in the case of large training dataset. Nevertheless, in the case of 50 times smaller training dataset, we suppose that the MC-CNN method cannot be successfully trained in contrast to the proposed method, which was successfully trained and demonstrated ACER=2.91% in this case. Additional experiments with heuristic strategies of feature improvement allowed us to decrease the classification error almost 1.84 times in comparison with the best basic method implementation.

5. ACKNOWLEDGMENTS

The reported study was funded by RFBR, project number 19-29-09045.

6. REFERENCES

[Abd21] Abdullakutty, F., Elyan, E., and Johnston, P. A review of state-of-the-art in Face Presentation Attack Detection: From early development to advanced deep learning and multi-modal fusion methods. Information fusion 75, pp. 55-69, 2021. DOI: 10.1016/j.inffus.2021.04.015.

[Anj11] Anjos, A., and Marcel, S. Counter-measures to photo attacks in face recognition: a public database and a baseline Biometrics (IJCB). 2011 international joint conference on Biometrics

(IJCB), pp. 1-7, 2011. DOI: 10.1109/IJCB.2011.6117503.

[Bha19] Bhattacharjee, S., Mohammadi, A., Anjos, A., and Marcel, S. (2019). Recent advances in face presentation attack detection. Handbook of Biometric Anti-Spoofing, pp. 207-228, 2019. DOI: 10.1007/978-3-319-92627-8_10.

[Bre19] Bresan, R., Pinto, A., Rocha, A., Beluzo, C., and Carvalho, T. (2019). Facespoof buster: a presentation attack detector based on intrinsic image properties and deep learning. arXiv preprint arXiv:1902.02845, 2019. 10.48550/arXiv.1902.02845.

[Chr00] Christianini, N. and Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press, 2000.

[Den21] Denisova, A. and Fedoseev, V. Presentation Attack Detection in Facial Authentication using Small Training Dataset Obtained by Multiple Devices. 2021 International Conference on Information Technology and Nanotechnology (ITNT), pp. 1-5, 2021. DOI: 10.1109/ITNT52450.2021.9649390.

[Du21] Du, Y., Qiao, T., Xu, M., and Zheng, N. (2021). Towards Face Presentation Attack Detection Based on Residual Color Texture Representation. Security and Communication Networks, pp. 6652727, 2021. DOI: 10.1155/2021/6652727.

[Ewa20] Ewald, K.E., Zeng, L., Zhengyao, Mawuli, C.B., Abubakar, H.S. and Victor, A. Applying CNN with Extracted Facial Patches using 3 Modalities to Detect 3D Face Spoof. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 216–221, 2020. DOI: 10.1109/ICCWAMTIP51612.2020.9317329.

[Fre12] de Freitas Pereira, T., Anjos, A., De Martino, J. M., and Marcel, S. LBP− TOP based countermeasure against face spoofing attacks. Asian Conference on Computer Vision, pp. 121-132, 2012. DOI: 10.1007/978-3-642-37410-4_11.

[Gal14] Galbally, J., Marcel, S. and Fierrez, J. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. IEEE transactions on image processing 23, No. 2, pp. 710–724, 2014. DOI: 10.1109/TIP.2013.2292332.

[Geo19] George, A., Mostaani, Z., Geissenbuhler, D., Nikisins, O., Anjos, A., and Marcel, S. Biometric face presentation attack detection with multi-channel convolutional neural network. IEEE Transactions on Information Forensics and Security 15, pp. 42-55, 2019. DOI: 10.1109/TIFS.2019.2916652.

[Har73] Haralick, R., Shanmugam, K., Dinstein, I. Textural features for image classification. IEEE TSMC 3, No. 6, pp. 610-621, 1973. DOI: 10.1109/TSMC.1973.4309314.

[Kow20] Kowalski, M. A Study on Presentation Attack Detection in Thermal Infrared. Sensors 20, No. 14, pp. 3988, 2020. DOI: 10.3390/s20143988.

[Kul12] Kulkarni, V.Y. and Sinha, P.K. Pruning of random forest classifiers: A survey and future directions. 2012 International Conference on Data Science & Engineering (ICDSE), pp. 64-68, 2012. DOI: 10.1109/ICDSE.2012.6282329.

[Liu09] Liu, C. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. Dissertation. Citeseer, 2009.

[Pen18] Peng, F., Qin, L. and Long, M. Face presentation attack detection using guided scale texture. Multimedia Tools and Applications 77, No. 7, pp. 8883-8909, 2018. DOI: 10.1007/s11042-017-4780-0.

[Pen20] Peng, F., Qin, L. and Long, M. Face presentation attack detection based on chromatic co-occurrence of local binary pattern and ensemble learning. Journal of Visual Communication and Image Representation 66, pp. 102746, 2020. DOI: 10.1016/j.jvcir.2019.102746.

[Wan19] Wang, G., Lan, C., Han, H., Shan, S., and Chen, X. Multi-modal face presentation attack detection via spatial and channel attentions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[Wan20] Wan, J., Guo, G., Escalera, S., Escalante, H. J., and Li, S. Z. Multi-Modal Face Presentation Attack Detection. Synthesis Lectures on Computer Vision 9, No. 1, pp. 1-88, 2020. DOI: 10.2200/S01032ED1V01Y202007COV017.

[Wen15] Wen, D., Han, H. and Jain, A. K. Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security 10, No. 4, pp. 746–761, 2015. DOI: 10.1109/TIFS.2015.2400395.

[Wu18] Wu, X., He, R., Sun, Z. and Tan T. A Light CNN for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security 13, No. 11, pp. 2884-2896, 2018. DOI: 10.1109/TIFS.2018.2833032.