Deep Color Spaces for Fingerphoto Presentation Attack Detection in Mobile Devices

Emanuela Marasco[1(✉)], Anudeep Vurity[1], and Asem Otham[2]

[1] Center for Secure and Information Systems,
George Mason University, Fairfax, USA
{emarasco,avurity}@gmu.edu
[2] Acuant, Los Angeles, USA
aothman@acuant.com

Abstract. Fingerphotos are fingerprint images acquired using a basic smartphone camera. Although significant progress has been made in matching fingerphotos, the security of these authentication mechanisms is challenged by presentation attacks (PAs). A presentation attack can subvert a biometric system by using simple tools such as a printout or a photograph displayed on a device. The goal of this research is to improve the performance of fingerphoto presentation attack detection (PAD) algorithms by exploring the effectiveness of deep representations derived from various color spaces. For each color space, different convolutional neural networks (CNNs) are trained and the most accurate is selected. The individual scores output by the selected CNNs are combined to yield the final decision. Experiments were carried out on the IIITD Smartphone Fingerphoto Database, and results demonstrate that integrating various color spaces, including the commonly used RGB, outperforms the existing fingerphoto PAD algorithms.

Keywords: Colorspace · Presentation attack detection · FingerPhoto

1 Introduction

As government services become more dependent on mobile technology, these devices become more subject to attacks that aim to gain unauthorized access to sensitive information. To address this issue, fingerprint identification has been deployed in mobile systems. The sensors embedded in mobile devices are small, and the resulting images acquired with them are, therefore, of limited size. The integration of common smartphones in the sensing and authentication process would enable relevant applications such as device unlocking and mobile payments [2]. However, while capacitive fingerprint sensors can be embedded in newer smartphones, billions of existing smartphones are being excluded. Furthermore, fingerprints are currently proprietary and inaccessible in smartphones.

B. Raman et al. (Eds.): CVIP 2021, CCIS 1567, pp. 351–362, 2022.
https://doi.org/10.1007/978-3-031-11346-8_31

Fingerphoto-based technologies can enable the use of fingerprints in smartphones for authentication on a large-scale such as National ID programs where accurate, usable, and low-cost systems are required. A fingerphoto is an image of the frontal part of a finger captured by a smartphone camera. Fingerphotos to fingerphotos matching has recently achieved a True Acceptance Rate (TAR) of 99.66% at False Acceptance Rate (FAR) of 0.1% by fusing four fingers, and a TAR of 85.62% at FAR = 0.1% by using individual fingers [2]. Accurate comparison of fingerphotos against slap fingerprints has also been achieved, with a TAR of 95.79% at FAR of 0.1% when fusing four fingers, and a TAR of 76.63% using individual fingers [2]. Despite relevant advances in matching, this technology is vulnerable to presentation attacks (PAs) [23]. PAs refer to techniques that inhibit the intended operation of a biometric capture system and interfere with the acquisition of the true identity [15]. PAs can either conceal an individual's identity or impersonate someone else. Presentation attack detection (PAD) modules classify biometric samples as being live (non-spoof) or fake (spoof). In this paper, we focus on two types of replica: *i)* print attacks, realized using a color paper-printout placed in front of the phone camera, and *ii)* photo attacks, carried out by displaying the original image in front of the capturing device.

Existing fingerphoto PADs based on the extraction of textural descriptors such as Dense Scale Invariant Feature Transform (DSIFT), Locally Uniform Comparison Image Descriptor (LUCID), and Local Binary Patterns (LBP) from RGB images did not reach high accuracy. Processing only RGB images may limit the approach since a spoof can only be modeled in terms of percentages of the three primaries composing its color in such a color model. Choosing an appropriate color space can provide a more robust analysis. Differences between live human fingers and display attacks can be detected better when they are modeled using suitable color descriptors. RGB may be ideal for image color generation but less suited for color description. In color models such as HSI (hue, saturation, intensity), the intensity component is decoupled from the color-carrying information (hue, saturation); thus, they are better aligned to how humans perceive color attributes.

The contribution of this paper is to explore deep representations of different color spaces and effectively integrate them for enhancing the performance of fingerphoto PADs. The proposed analysis includes robustness to unconstrained acquisition including background variations. The rest of the paper is structured as follows: Sect. 2 reviews research conducted on fingerphoto PADs, Sect. 3 discusses the proposed approach, Sect. 4 presents the experimental results, while Sect. 5 draws our conclusions and discusses future work.

2 Related Work

In fingerprint scanners, PAs have been detected by either gathering further evidence of the vitality of the subject (e.g., sensing blood circulation, or fluids - perspiration patterns - secreted when touching surfaces) or by passive methods

detecting the presence of known materials (e.g., material structure, lack of high-resolution detail) [15]. Several software-based methods, including Fourier Transform (FT), Local Binary Patterns (LBP), Binarized Statistical Image Features (BSIF), Local Phase Quantization (LPQ), Weber Local Image Descriptor or Histograms of Invariant Gradients (HIG), have been investigated for PAD [5,6]. Existing efforts in biometric liveness detection have been expanded by considering the assessment of activities, motivations, intent, and capabilities of attackers. However, these liveness detection approaches are not explicitly designed for mobiles and generally unsuited for portable devices [1].

At the same time, the possibility of spoofing fingerphoto-based systems is real, and despite the risk, only a few research efforts have been spent to mitigate the issue. In 2013, Stein *et al.* [22] discussed a technique for fingerphoto PAD that measures the light reflected from the finger exposed to the LED of the camera, as well as the position, distance, and sharpness properties of the finger. An overall Equal Error Rate (EER) of 3% was reported on video stream data collected from 37 subjects. In 2016, Taneja *et al.* evaluated LBP, DSIFT, and LUCID on a database of spoofs that they created by extending the previously published IIITD Smartphone Fingerphoto Database [23]. The lowest EER reported is 3.7% and was achieved by SVM trained with LBP features on the complete dataset. DSIFT and LUCID reported EER of 5.37 and 22.22%, respectively. Although the EER was not very high, the performance was poor when considering TAR at FAR = 0.1%. In general, descriptors commonly used in spoofing literature yield very poor results for fingerphotos.

In 2018, Wasnik *et al.* discussed an approach in which input images are processed at multiple scales through a Frangi filter. From the generated maximum filter response (MFR) images, LBP, HOG, and BSIF features are extracted [24]. Bona Fide Presentation Classification Error Rate (BPCER) was 1.8% for print photo attacks, 0.0% for display attacks, and 0.66% for replay attacks at Attack Presentation Classification Error Rate (APCER) = 10% by a SVM. Results pertain to data collected from 50 subjects using iPhone and iPad Pro devices. Fujio *et al.* investigated fingerphoto spoof detection under noisy conditions (e.g., blurring). The original images were filtered to simulate distortions due to camera defocus and hand movements effects. An AlexNet was trained using the database created by Taneja *et al.,* and a Half Total Error Rate (HTER) of 0.04% was reported [4,23]. This model showed robustness to blurred images.

3 The Proposed System

To date, fingerphotos have been processed only in the form of RGB images without modifying them for classification tasks including PAD. In this paper, we discuss a framework in which fingerphoto RGB images are converted into multiple color spaces before classification.

Color is a powerful descriptor that can simplify object identification. A color space is a specific organization of colors helping to produce a digital representation of colors [7]. In a color space, each color is represented by a single point which

provides a three-dimensional object containing all realizable color combinations. Radiance, luminance, and brightness are the basic quantities used to describe the quality of a chromatic light source. Based on experimental evidence, 65% of the human cones are sensitive to human light, 33% to green light, and only 2% to blue. Brightness embodies the chromatic notion of intensity; luminance indicates the amount of energy from a light source perceived by an observer, while radiance indicates the total amount of energy that flows from the light source. In image processing, color models can be divided into three main categories: *i)* device-oriented, where color is specified in a way that is compatible with the hardware tools used; *ii)* user-oriented, utilized to link human operators to the hardware used (human perception of color); and *iii)* device-independent, where color signals do not depend on the characteristics of the given device which allows the connection of different hardware platforms.

In the proposed system, various architectures of Convolutional Neural Networks (CNNs) are trained to classify images in different color spaces [7]. Each network is fed with raw inputs (e.g., normalized images), which transform into gradually higher levels of representation (e.g., edges, local shapes, object parts), and the first few layers act as a feature extractor [13]. In convolutional networks, three architectural ideas ensure shift and distortion invariance: local receptive fields, shared or repeated weights and spatial (or temporal) sub-sampling [14]. The capacity of CNNs can be controlled by varying their depth and breadth; furthermore, valid assumptions about the nature of the images can be made (e.g., locality of pixel dependencies) [13]. Thus, CNNs have much fewer connections than the standard forward-feed neural networks with layers of similar size, resulting in easier training. As illustrated in Fig. 1, for a given color space, we first determine which CNN classifies it with the highest accuracy. Then, the individual confidence scores yielded by the most accurate networks are integrated at score-level to output the overall classification decision.

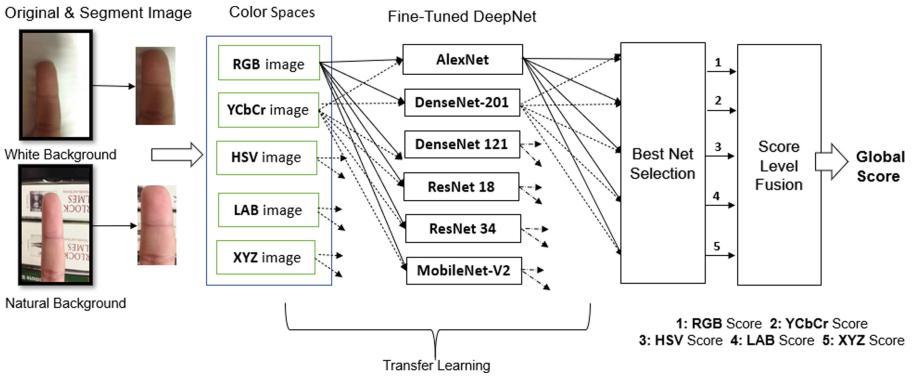


Fig. 1. The Proposed Architecture. The input RGB fingerphoto image is segmented and then converted into various color spaces. Each individual network learns to classify a single color model. For each color space, the most accurate deep net is selected. The contributions from the individual color spaces are then combined to obtain the final decision.

Information fusion in biometrics has been proven to be able to enhance recognition accuracy compared to a system that exploits only a single source. An effective fusion scheme must be implemented to integrate the evidence provided by multiple sources [11]. Biometric information can be combined at sensor or feature-level, or decision or score-level. While the amount of information to integrate progressively decreases from the sensor-level to the decision-level, the degree of noise also decreases [16]. Different integration schemes can be designed; the most effective are explained as follows:

- Feature-level Fusion. Features extracted from different color spaces are concatenated to create a single feature vector that can feed a classifier.
- Score-level Fusion. The combination of different color spaces occurs after obtaining a deep representation of each individual color space [18]. This scheme combined the confidence scores output by the individual CNNs. A threshold determined in training is then applied to the score produced by the fusion scheme.
- Decision-level Fusion. Individual classification decisions are fused. This approach is bandwidth efficient since only decisions requiring a single bit, are transmitted to the fusion engine.

In this work, we implement score-level fusion in which the scores output by the individual deep networks are fused via simple sum rule.

Before training the models, the images were segmented for background removal. The segmentation algorithm uses an adaptive skin color thresholding [20]. The procedure converts them into RGB image into the corresponding CMYK scale in which the magenta component is thresholded using the Otsu method to generate the binary mask representing the skin region of the fingerphoto [17]. A rectangular ROI is then determined and cropped. The segmented images are then transformed into different color spaces. Before training, the segmented images are resized by preserving the aspect ratio, see Table 1.

3.1 Architecture and Fine-Tuning of the DeepNets

Given the limited availability of fingerphoto spoof data, fine-tuning was applied for implementing transfer learning [10]. To adapt the pre-trained models to the PAD binary classification task, the last fully-connected layer was modified from 1000-way SoftMax to 2-way SoftMax. The parameters were estimated by using the Adaptive Moment Estimation (Adam) Optimization algorithm. Adam optimizer computes an adaptive learning rate for each parameter, while gradient descent uses a constant learning rate. Optimizers update the model in response to the output of the loss function, and thus they assist in minimizing it. Let $\mathbf{Y}$ be the vector of labels (i.e., the ground truth) and $\mathbf{Y'}$ be the vector of the predictions. The Softmax layer is combined with the Cross-Entropy (CE) Loss. The loss represents the distance between $\mathbf{Y}$ and $\mathbf{Y'}$, see Eq. (1). It increases as the predicted probability diverges from the actual label. The smaller the loss, the better is the model.

$$\text{CELoss}: \quad D(\mathbf{Y}, \mathbf{Y'}) = -\frac{1}{N}\sum Y_i log(Y_i') \tag{1}$$

where N is the number of training samples.

- **AlexNet**: The architecture is featured by five convolutional and three fully-connected layers, with ReLU activation function. The architecture is featured by Rectified Linear Units (ReLUs) with activation $f = max(x, 0)$ for their ability to be trained faster than *tanh* units, with a local response normalization applied. The model was pre-trained on ImageNet, and the last fully-connected layer was fed to a 1000-way Softmax [12].
- **ResNet**: The Deep Residual learning framework aims to address the problem of performance degradation in deeper networks. A residual network (ResNet) results from constructing a deeper counterpart of a shallower architecture in which the added layers are identity mapping while the other layers are copied from the learned shallower model. The stacked layers fit a residual mapping $F(x) = H(x) - x$ where $H(x)$ is the desired underlying mapping that can be recast into $F(x) + x$ [8]. The residual mapping is assumed to be easier to be optimized than the original one. Depending on the number of layers, there exist different variants of ResNet such as ResNet 18, ResNet 34, ResNet 121, and ResNet 201.
- **MobileNet-V2**: The architecture is based on an inverted residual structure in which the input and output of the residual block are thin bottleneck layers. The convolution is split into two separate layers: depthwise convolution that performs lightweight filtering by applying a single convolutional filter per input channel, and pointwise convolution that builds new features through computing linear combinations of the input channels [19]. Depthwise convolution reduces computation compared to traditional layers.
- **DenseNet**: In a DenseNet, each layer is connected to every other layer in a feed-forward fashion. This network utilizes dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other [9]. The deep layers can access all the feature-maps created by preceding levels, thus encouraging features reused. DenseNet variants include DenseNet 121, DenseNet 161, Dense 201.

3.2 Implementation of Mobile Deep Learning

Despite the relative efficiency of the CNNs local architecture, the training of large CNNs needs to be facilitated by powerful GPUs and highly-optimized implementation of 2D convolution applied in large-scale to high-resolution images.

Low-power consumption is one of the main factors driving the development of mobile processors. Snapdragon is a family of mobile systems on a chip (SoC) processor architecture provided by Qualcomm [21]. Qualcomm Snapdragon SoC is built around the Krait processor architecture. Adreno GPU in this architecture delivers improved advanced graphics performance. Specifically, Snapdragon 800 processor consists of the 28nm HPm quad-core Krait 400 CPU for high performance, Adreno 330 GPU for improved graphics performance, Hexagon DSP for low power operation, and Gobi True 4G LTE modem for connectivity. To support the next-generation data-centric mobile devices, processor architectures

must be designed considering these approaches [3]. In Table 1, we report the memory and computational power required to execute the algorithms used in this work. The CNNs were trained using NVIDIA k80 GPU (12 GB RAM per GPU). The notation FLOPs indicates floating point operations per second.

Table 1. Memory and computational power requirements

	Input size	Parameter memory	Feature memory	FLOPs
AlexNet	227 × 227	233 MB	3 MB	727 MFLOPs
ResNet18	224 × 224	45 MB	23 MB	2 GFLOPs
ResNet34	224 × 224	83 MB	35 MB	4 GFLOPs
DenseNet 121	224 × 224	31 MB	126 MB	3 GFLOPs
DenseNet 201	224 × 224	77 MB	196 MB	4 GFLOPs
MobileNet	224 × 224	17 MB	38 MB	579 MFLOPs

4 Experimental Results

4.1 Dataset

For this study, we used the IIITD smartphone fingerphoto database of live images and the spoof fingerphotos created from it [20]. Samples pertain to 64 individuals in two different backgrounds, in controlled and uncontrolled illumination. Two subsets White-Indoor (WI) and White-Outdoor (WO), are created by capturing fingerphotos with white background in both indoor (controlled illumination) and outdoor (uncontrolled illuminations) conditions. Similarly, the capture with a natural background in both indoor and outdoor conditions generated two subsets Natural-Indoor (NI) and Natural-Outdoor (NO). Each subset NI, NO, WI, and WO contains 8 samples of the right index and right middle fingers per individual, therefore a total of 1024 images (64 subjects x 2 fingers x 8 instances).

The spoof fingerphotos were created from the IIITD smartphone fingerphoto database by randomly selecting 2 instances out of 8 per subject [23]. Three photo attacks using three different mobile devices (Apple iPad, Dell Inspiron laptop, and Nexus) as well as one printout attack (HP Color-LaserJet CP2020 Series PCL6 printer at 600 ppi) were created. OnePlusOne and Nokia devices were used during capture under spoof attacks. The display mechanisms were realized using: (a) Apple iPad with Retina display with 2048 × 1536 resolution, (b) Nexus 4 with 1280 × 760 resolution, and (c) Dell Inspiron N5110 Laptop with 1280 × 720 resolution. Spoof data pertains to 64 subjects and featured by 2 illumination types, 2 background variations, 2 finger instances, 2 capture mechanisms and 4 display mechanisms (1 printout + 3 photos) for a total of 8192 spoofs. The complete database used in this study consists of 12,288 fingerphoto images (4096 live and 8192 spoof).

4.2 Evaluation Procedure

To assess the proposed framework, we refer to the performance metrics defined in the ISO/IEC 30107-3 standard on biometric presentation attack detection part related to testing and reporting performance metrics for evaluating biometric presentation attacks. The assessment scheme is reported below:

- Attack Presentation Classification Error Rate (APCER): Proportion of attack presentations incorrectly classified as normal presentations, i.e., false acceptance of spoof samples.
- Normal Presentation Classification Error Rate (NPCER): Proportion of normal presentations incorrectly classified as attack presentations, i.e., false rejection of live samples.
- Equal Error Rate (EER): The intersection point of the percentage of normal presentation classification error rate and attack presentation classification error rate.
- Receiver Operating Characteristic (ROC) curves to assess the accuracy.

In this paper, we establish a baseline in the two scenarios white vs. white and natural vs. natural background; then, we study the effects of background changes without the influence of variations in lighting conditions (indoor vs. outdoor). We also analyze robustness versus both illumination and background changes by training the system on the complete database. In all the experiments, data was split into 50% training, and 50% testing and the subjects were mutually exclusive between training and testing.

4.3 Results

In this section, we discuss the performance of the proposed approach and compare it to existing algorithms used for fingerphoto spoof detection to date. From the histograms shown in Fig. 2, we can observe how good is the separation between live and spoof in the HSV model. In this color space, the characteristics hue (H), saturation (S), and brightness (B or V) are generally used to distinguish one color from another. The H attribute represents the dominant color, S is the level of purity or amount of white light mixed with a hue, and V indicates intensity. In the H channel, live distributions are concentrated below 50, which may be typical of actual live skin color. In the S channel, live samples can reach noticeable peaks in the ranges 100–130 and 50, which differs from spoof samples. In V, only live images are distributed above 150, which indicates high brightness.

Table 2 reports the results of AlexNet fine-tuned on individual color spaces under background variations. The experiments were conducted using different display and capture mechanisms. XYZ reached the lowest EER in the white vs. natural setting, which shows promising performance whether the training is carried out on images with white background and the authentication in more unconstrained conditions. XYZ is efficient on low-resolution capture mechanisms with natural background. When the complete dataset is used, HSV achieved

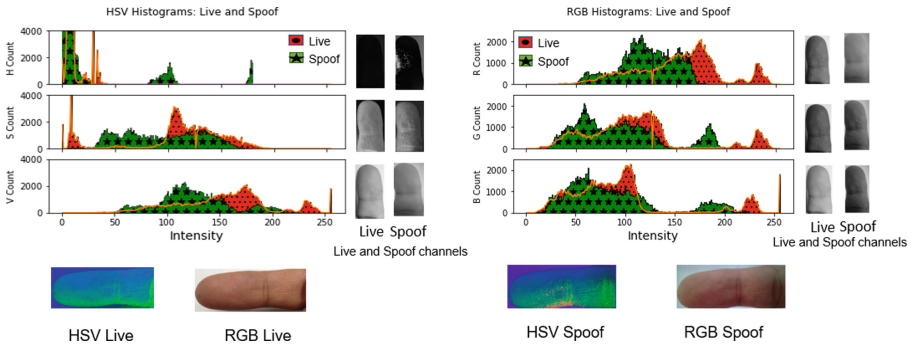


Fig. 2. Histograms of the RGB and HSV color spaces. The y-axis refers to the number of pixels for each component H, S, V, and R, G, B.

Table 2. EER % for robustness to background variations

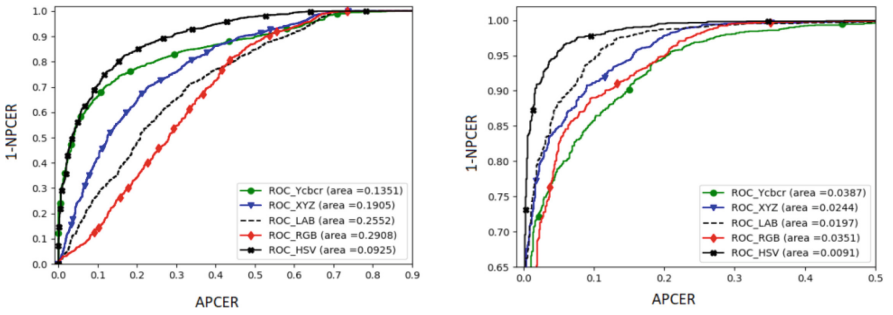
		White vs. White					Natural vs. Natural				
Display	Capture	RGB	YCbCr	HSV	LAB	XYZ	RGB	YCbCr	HSV	LAB	XYZ
Print	Nokia	5.712	5.314	1.953	6.250	4.687	3.154	5.078	**1.954**	5.859	2.734
	OPO	2.734	7.421	3.320	5.957	3.320	4.351	6.641	3.515	4.687	3.955
iPad	Nokia	2.978	3.050	1.562	4.687	4.296	4.296	5.517	2.343	5.078	3.906
	OPO	3.467	5.851	2.734	4.882	3.906	5.371	6.641	3.515	4.785	4.248
Smart Phone	Nokia	4.541	3.185	2.246	4.394	2.343	3.955	4.687	2.734	4.687	3.166
	OPO	0.051	7.212	3.906	5.761	3.222	3.515	4.638	**1.953**	4.980	3.515
Laptop	Nokia	2.099	4.003	1.562	4.296	3.150	**1.953**	5.469	2.050	5.419	3.102
	OPO	**0.035**	4.296	1.367	4.199	3.906	3.417	3.516	**1.953**	5.859	4.736
Complete Dataset		3.112	4.980	1.953	4.199	4.492	3.710	3.516	3.417	5.273	4.002

		White vs. Natural					Natural vs. White				
Display	Capture	RGB	YCbCr	**HSV**	LAB	**XYZ**	RGB	YCbCr	HSV	LAB	XYZ
Print	Nokia	37.540	22.656	14.063	33.594	24.805	17.188	17.188	9.766	16.016	14.063
	OPO	35.938	16.058	19.141	37.109	28.418	10.938	12.012	5.859	11.523	12.500
vsiPad	Nokia	6.641	12.891	6.738	34.765	**4.004**	**3.516**	7.813	1.953	6.543	4.590
	OPO	8.789	14.844	14.746	14.746	9.082	5.859	8.789	4.297	11.328	10.547
Smart Phone	Nokia	10.156	**4.541**	8.398	14.063	12.109	**3.187**	7.031	4.688	7.227	6.641
	OPO	7.910	10.938	8.984	13.574	9.473	5.762	10.938	3.906	9.668	5.859
Laptop	Nokia	7.813	**4.102**	10.547	17.188	9.668	5.469	8.594	**3.129**	8.594	6.641
	OPO	12.891	9.375	**4.688**	15.430	8.203	5.762	8.984	5.176	8.887	5.957
Complete Dataset		34.668	20.215	17.871	36.645	27.246	10.156	13.184	5.176	7.813	9.473

the lowest EER. In this scenario, multiple PAs types are mixed and variations pertain to both illumination (outdoor/indoor) and background (white/natural). The model trained on RGB images performed well on attacks realized with laptop-opo and smartphone-opo. High-resolution capture mechanisms seem to have a positive impact on the RGB color model when the background is white. ROC curves pertaining to the analysis of the robustness of each color model to

background variations are illustrated in Fig. 3. This specific case study refers to the use of AlexNet. In cross-background, while the performance of most color models including RGB significantly deteriorates, HSV exhibits robustness. This represents a promising performance when dealing with unconstrained acquisition.

For each color space, we selected the most accurate deep network. From Table 3, we can observe that on the complete dataset the lowest EER is provided by ResNet-34 for the color spaces RGB and XYZ, MobileNet for HSV and YCbCr; while ResNet-18 represents the best architecture for the LAB color space. Table 4 reports the fusion results of the proposed framework and compares it to existing fingerphoto PADs. Color spaces were combined by averaging the scores output by the individual best networks. The EER obtained by the fusion of the best three color scores is the lowest: 2.12%. RHY indicates the model fusing RGB, HSV, and YCbCr; RHYL fuses RGB, HSV, YCbCr and LAB; RHYLX fuses the five color spaces used in this work. RHY provides the best performance.



(a) White vs. Natural (b) Natural vs. White

Fig. 3. ROC curves for the experiments on the robustness to background variations.

Table 3. Selection of the best networks on the complete dataset (EER%)

	Complete dataset				
	RGB	HSV	Ycbcr	LAB	XYZ
MobileNet- V2	3.541	**2.707**	**3.445**	**4.21**	4.871
DenseNet-201	14.258	15.36	15.258	15.266	17.587
ResNet-34	**3.174**	2.714	4.207	4.612	**4.772**
ResNet-18	3.242	3.112	4.052	4.124	4.854
DenseNet 121	9.981	8.127	12.211	11.948	14.571
AlexNet	3.361	2.811	4.505	5.571	4.829

Table 4. Comparison of the proposed system to the state-of-the-art (EER%)

	RHY	RHYL	RHYLX	LBP +SVM	DSIFT +SVM	LUCID+SVM
Complete	2.12%	2.71%	3.291%	3.71%	5.37%	2.22%

5 Conclusions

Fingerphotos acquired with common smartphone cameras could enable an effective approach for authentication. Since this technology is vulnerable to spoof attacks, the proposed efforts focus on enhancing its security. We discuss the first investigation that integrates multiple color models for accurate fingerphoto PADs. Our methodology transforms fingerphoto RGB images into various color spaces and trains a different deep network for each of them. For each color space, the best deep representation is determined and the corresponding individual outputs are combined at the score-level. Experiments were carried out using different display mechanisms on a publicly available database. Results demonstrate the superiority of the proposed framework compared to existing approaches that mainly operate and process only RGB images. We will extend the experiments by exploring additional integration strategies including deep fusion.

Acknowledgement. This work was supported by the National Science Foundation (NSF) under award CNS-1822094.

References

1. Akhtar, Z., Micheloni, C., Piciarelli, C., Foresti, G.L.: MoBio_LivDet: mobile biometric liveness detection. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 187–192 (2014)
2. Deb, D., Chugh, T., Engelsma, J., Cao, K., Nain, N., Kendall, J., Jain, A.: Matching fingerphotos to slap fingerprint images. arXiv preprint arXiv:1804.08122 (2018)
3. Deng, Y.: Deep learning on mobile devices: a review. In: Mobile Multimedia/Image Processing, Security, and Applications 2019, vol. 10993, pp. 52–66. SPIE, May 2019
4. Fujio, M., Kaga, Y., Murakami, T., Ohki, T., Takahashi, K.: Face/fingerphoto spoof detection under noisy conditions by using deep convolutional neural network. In: BIOSIGNALS, pp. 54–62 (2018)
5. Ghiani, L., Hadid, A., Marcialis, G.L., Roli, F.: Fingerprint liveness detection using binarized statistical image features. In: IEEE Biometrics: Theory, Applications and Systems (BTAS), pp. 1–6 (2013)
6. Gottschlich, C., Marasco, E., Yang, A., Cukic, B.: Fingerprint liveness detection based on histograms of invariant gradients. In: IEEE International Joint Conference on Biometrics, pp. 1–7 (2014)
7. Gowda, S.N., Yuan, C.: ColorNet: investigating the importance of color spaces for image classification. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11364, pp. 581–596. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20870-7_36
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition, pp. 770–778 (2016)

9. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks, pp. 4700–4708 (2017)
10. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems (2014)
11. Kittler, J., Hatef, M., Duin, R.P., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **20**(3), 226–239 (1998)
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**, 1097–1105 (2012)
13. LeCun, Y., Huang, F., Bottou, L.: Learning methods for generic object recognition with invariance to pose and lighting. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, p. II-104 (2004)
14. LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time series. In: The Handbook of Brain Theory and Neural Networks, vol. 3361, no. 10 (1995)
15. Marasco, E., Ross, A.: A survey on antispoofing schemes for fingerprint recognition systems. ACM Comput. Surv. **47**(2), 28:1–28:36 (2014). https://doi.org/10.1145/2617756. http://doi.acm.org/10.1145/2617756
16. Marasco, E., Ross, A., Sansone, C.: Predicting identification errors in a multibiometric system based on ranks and scores. In: Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems, September 2010
17. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
18. Ross, A., Jain, A., Nandakumar, K.: Information fusion in biometrics. In: Handbook of Multibiometrics, pp. 37–58 (2006)
19. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2: Inverted residuals and linear bottlenecks, pp. 4510–4520 (2018)
20. Sankaran, A., Malhotra, A., Mittal, A., Vatsa, M., Singh, R.: On smartphone camera based fingerphoto authentication. In: IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–7 (2015)
21. Singh, M.P., Jain, M.K.: Evolution of processor architecture in mobile phones. Int. J. Comput. Appl. **90**(4), 34–39 (2014)
22. Stein, C., Nickel, C., Busch, C.: Fingerphoto recognition with smartphone cameras, pp. 1–12 (2012)
23. Taneja, A., Tayal, A., Malhorta, A., Sankaran, A., Vatsa, M., Singh, R.: Fingerphoto spoofing in mobile devices: a preliminary study. In: IEEE Biometrics: Theory, Applications and Systems (BTAS), pp. 1–7 (2016)
24. Wasnik, P., Ramachandra, R., Raja, K., Busch, C.: Presentation attack detection for smartphone based fingerphoto recognition using second order local structures. In: 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 241–246 (2018)