Cross-illumination Evaluation of Hand Crafted and Deep Features for Fusion of Selfie Face and Ocular Biometrics

Leena Kondapi[1], Ajita Rattani[2] and Reza Derakhshani[1]
[1]Department of Computer Science and Electrical Engineering
University of Missouri at Kansas City
Email: lkgp7@mail.umkc.edu
[2]Department of Electrical Engineering and Computer Science
Wichita State University
Wichita, Kansas
Email: ajita.rattani@wichita.edu

***Abstract*—This paper addresses the implementation of a multi-unit biometric system. Results are shown for multi-unit classification with VISible light mobile Ocular Biometric (VISOB) dataset using feature descriptors such as Local Binary Patterns (LBP) and Histogram of oriented gradients (HOG). We also evaluate the pre-trained deep learning models such as VGG16, ResNet18, MobileNetV1, MobileNetV2, and LightCNN9. Experimental evaluation on large scale VISOB dataset suggests that feature-level fusion followed by score-level fusion of left ocular region, right ocular region and face region in office light condition, daylight and dims condition has provided Equal Error Rates (EER) of 9.3%, 8.0% and 10.6% respectively. Also, combining the pre-trained models using feature fusion decreased the EER even further.**

I. Introduction

Biometric authentication on mobile devices has become a convenient and important means of verifying identities for secured digital access and services such as device-unlock mobile banking [1]. In this context, face and ocular biometrics in visible spectrum have received increased attention from research community and industry alike. One main reason is their natural and low- effort user experience, similar to taking a selfie of ones face or ocular regions, for mobile use case [1],[2]. Textural descriptors, image key points, and patch descriptors (such as BSIF, LBP and HOG) as well as several pre-trained deep learning models (such as VGG16, ResNet18, MobileNetV1, and MobileNetV2) have been mostly used either in a learning or non-learning- based framework for identity verification using these modalities [3],[4].

Nonetheless, proper integration of biometric in mobile application has its own challenges. This is because of degradation in acquired biometric samples due to factors such as illumination variation, makeup, specular reflection, defocus and motion blur. Many of these factors arise from device mobility and operation in uncontrolled environments. Under such conditions, biometric test samples may differ significantly from the ones acquired during enrollment (reference template). Consequently, the recognition performance is usually lower, especially for non-training-based systems [4] (unless the development set is small). Figure 1 shows examples of ocular images acquired using the front facing camera of an iPhone 5s.

One of the possible ways to mitigate the effect of the aforementioned factors is multi biometrics, or more generally employing and fusing multiple sources of information for performance enhancement. These various sources of information can be multiple modalities (such as fusing face and fingerprints), multiple classifiers (such as Support Vector Machines and Neural networks, possibly operating on the same source), and multiple feature representations (such as LBP and BSIF), at various fusion levels [2].

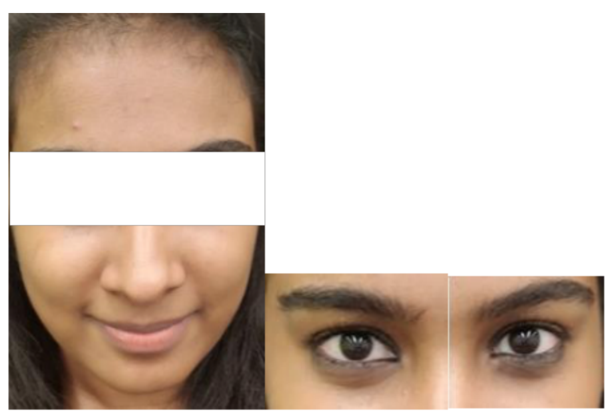

Fig. 1. Example of selfie face and ocular regions captured with an iPhone 5s.

The aim of this paper is performance enhancement using fusion of selfie face images and ocular region using hand crafted and deep features, i.e. a combination of the above-mentioned fusion paths.

II. Proposed Methodology

A. Handcrafted Features

The following hand crafted features, which are widely used for object detection were used to develop our multi-modal biometric system:

1) Local Binary Patterns (LBP): LBP is a simple yet efficient texture operator which labels pixels of an image by thresholding the neighborhood of each pixel and using the result as a binary number.
2) Histogram of Oriented Gradients (HOG): HOG decomposes an image into small square cells, computes the histogram of oriented gradients in each cell, normalizes the result using a block-wise pattern, and return a descriptor for each cell.

Using handcrafted features, namely LBP and HOG, we develop a multi-modal biometric system that fuses face and ocular modalities acquired from a mobile device at feature as well as score levels. In feature level fusion, the features from multiple sources are concatenated together followed by dimensionality reduction. In score level fusion, match scores of different biometric matchers are consolidated together in order to arrive at a better final recognition decision [2].

To this aim, the proposed system extracts LBP and HOG descriptors from left and right ocular regions and face using smartphones. The extracted LBP and HOG descriptors are concatenated together and normalized to a common range. This is followed by dimensionality reduction using a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). PCA and LDA linearly project the feature vectors such that the new, more compact set of k dimensions preserve variance or classifiability of the original d dimensions [5]. The Euclidean distances, between reduced feature vectors associated to reference template and test images are used as the similarity score for each modality.

B. Pre-trained models

Several pre-trained deep learning models were tested on VISOB dataset on left and right ocular regions, as well as face images. Areas under ROCs and EERs were used as measures of success. We found EERs when matching images acquired under different lighting conditions. We used multiple pre-trained deep learning networks, namely VGG16, ResNet18, LightCNN9, MobileNetV1 and MobileNetV2, as well as a ResNet-like model[1] developed using VISOB dataset to extract the features from the deep learning models to minimize the error rate and maximize the decidability and the ROC Area Under the Curve (AUC).

The following pre-trained models were tested for left ocular region, right ocular region and the face:

1) VGG16: VGG16 (also called Oxford Net) is a convolutional neural network architecture named after its developer, Visual Geometry Group. VGG network is characterized by its simplicity, using only 33 convolutional layers stacked on top of each other. Reducing the size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a SoftMax classifier.
2) ResNet18: Here the network layers are reformulated as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. These residual networks are easier to optimize and can gain accuracy from considerably increased depth.
3) MobileNetV1 and MobileNetV2: MobileNets are one of the more efficient models developed for mobile and embedded vision applications. MobileNets are based on a streamlined architecture.
4) LightCNN9: This model is a framework for learning compact embeddings from large-scale face data with noisy labels. A variation of max out activation, called Max-Feature-Map (MFM), is introduced into each convolutional layer.
5) ResNet-like model[1]: This model was trained using Adam optimizer [12] with the learning rate of lr = 0.001 for 150 epochs and with the batch size of 32 on VISOB dataset using SoftMax with categorical cross-entropy as the loss function. The ResNet18 pre-trained model was fine- tuned for ocular recognition using transfer learning. The transfer learning process is divided into two stages as follows:
    a) Stage 1: All the layers except the final classification layer (consisting of 400 outputs) are set as non-trainable and the whole network was trained for 10 epochs. We used Adam optimizer with lr = 0.001 and with batch size of 32.
    b) Stage 2: In this stage, all the layers were set as trainable and the models were trained for another 20 epochs with a reduced learning rate of lr = 0.0001.

This two-stage transfer learning process ensures that the newly initialized classification layer will not distort model weights too quickly and drastically, as they are considered to be at desirable ranges.

The next section provides the experimental validation and results for the above methods, followed by conclusion and future work.

III. Experimental Validation

A. Dataset

The dataset considered in this study, further described in [12], is the VISible light mobile Ocular Biometric (VISOB) dataset of ICIP2016 (challenge version). VISOB is publicly available and consists of eye images from 550 healthy adult volunteers acquired using three different smartphones i.e., iPhone 5s, Samsung Note 4 and Oppo N1.

B. Experimental protocol

1) The simple yet proven average score level fusion (sum rule) using LBP and HOG feature descriptors was used for fusion of match scores from left and right ocular regions, as well as the selfie faces. Our experimental

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

analysis is conducted over VISOB dataset. Our results proved the efficacy of the proposed multimodal fusion for smartphones. We first extract the features from the left and right ocular region followed by the face region and concatenate them together. Feature dimensionality is reduced using PCA and LDA. The scores from the output are fused using the sum rule at score level.

2) Using the pre-trained models, on face and ocular regions, the features from VGG16, ResNet18 and ResNet-like models are fused for left and ocular regions, and the features from VGG16, ResNet18, MobileNet, and LightCNN9 are fused for the face. The experimental results show that the EER reduced after feature fusion. The following experimental protocol was followed:
   a) The enrollment and test images were drawn from the similar lighting conditions (office, daylight, and dim lighting conditions).
   b) The images for enrollment came from office lighting condition and test from daylight and dim lighting conditions.

C. Results

This section discusses and compares the experimental results from the above mentioned two methodologies separately, namely score fusion of hand crafted features and feature fusion using pre-trained deep learning models.

TABLE I
EERS AND AUCS FROM SCORE LEVEL FUSION OF FACE AND OCULAR REGIONS (LEFT AND RIGHT) USING LBP AND HOG FEATURES ACQUIRED IN OFFICE LIGHT CONDITION

Method	AUC[0,1]	EER
Left ocular region	0.92	15.7
Right ocular region	0.91	16.8
Selfie face	0.94	12.5
Proposed fusion-based system	0.95	9.3

Table I shows testing EER and AUC values for enrolment and test images acquired in office light conditions. It can be seen that the proposed fusion based system reduces the EER from 16.8% (Right Ocular) to 9.3%.

TABLE II
EER AND AUCS FROM SCORE LEVEL FUSION OF FACE AND OCULAR REGIONS (LEFT AND RIGHT) USING LBP AND HOG FEATURES ACQUIRED IN DAYLIGHT CONDITION

Method	AUC[0,1]	EER
Left ocular region	0. 93	12.5
Right ocular region	0. 93	13.9
Selfie face	0.96	9.2
Proposed fusion-based system	0.96	8.0

Table II shows testing EER and AUC values for enrolment and test images acquired in daylight conditions. It can be seen that the proposed fusion-based system reduces the EER from 13.9% (Right Ocular) to 8.0%.

TABLE III
EER AND AUCS FROM SCORE LEVEL FUSION OF FACE AND OCULAR REGIONS (LEFT AND RIGHT) USING LBP AND HOG FEATURES ACQUIRED IN DIM CONDITION

Method	AUC[0,1]	EER
Left ocular region	0. 91	17.4
Right ocular region	0. 89	18.2
Selfie face	0.94	11.3
Proposed fusion-based system	0.95	10.6

Table III shows EER and AUC values for enrollment and test images acquired in dim lighting conditions. It can be seen that the proposed fusion-based system reduces the EER from 18.2% (Right Ocular) to 10.6%.

TABLE IV
EERS WHEN MATCHING FACES ACROSS DIFFERENT LIGHTING CONDITIONS USING PRE-TRAINED DEEP LEARNING MODELS

Model	Enrollment	Verification	EER for face region(%)
VGG16	Office	Office	18.88
	Office	Daylight	28.30
	Office	Dim	31.83
	Daylight	Daylight	18.38
	Dim	Dim	14.83
ResNet18	Office	Office	14.14
	Office	Daylight	23.36
	Office	Dim	27.96
	Daylight	Daylight	14.80
	Dim	Dim	9.37
MobileNetV1 05	Office	Office	14.12
	Office	Daylight	26.87
	Office	Dim	29.80
	Daylight	Daylight	13.35
	Dim	Dim	9.87
MobileNetV1 10	Office	Office	11.83
	Office	Daylight	20.72
	Office	Dim	25.26
	Daylight	Daylight	11.66
	Dim	Dim	7.06
MobileNetV2 05	Office	Office	12.50
	Office	Daylight	21.47
	Office	Dim	22.02
	Daylight	Daylight	14.61
	Dim	Dim	9
MobileNetV2 10	Office	Office	14.04
	Office	Daylight	22.58
	Office	Dim	27.65
	Daylight	Daylight	16.11
	Dim	Dim	9.41
LightCNN9	Office	Office	3.55
	Office	Daylight	5.60
	Office	Dim	6.09
	Daylight	Daylight	2.87
	Dim	Dim	1.68

Table IV show the EERs for face. As mentioned in the introduction, they are measure at different lighting conditions.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

TABLE V
EERS WHEN MATCHING ACROSS DIFFERENT LIGHTING CONDITIONS FOR LEFT AND RIGHT OCULAR REGIONS USING PRE-TRAINED DEEP LEARNING MODELS

Model	Enrollment	Verification	EER for left ocular region(%)	EER for right ocular region(%)
VGG16	Office	Office	23.33	24.22
	Office	Daylight	40.18	38.14
	Office	Dim	41.31	41.51
	Daylight	Daylight	25.94	27.16
	Dim	Dim	24.85	25.37
ResNet18	Office	Office	19.74	19.73
	Office	Daylight	37	36.61
	Office	Dim	37.85	36.98
	Daylight	Daylight	23.97	24.01
	Dim	Dim	37.85	36.98
ResNet-like model	Office	Office	23.33	24.22
	Office	Daylight	40.18	38.14
	Office	Dim	41.31	41.51
	Daylight	Daylight	25.94	27.16
	Dim	Dim	24.85	25.37

Table V shows the EERs for ocular regions. As mentioned in the introduction, they are reflect matching under varying lighting conditions.

TABLE VI
EERS FOR LEFT OCULAR REGION AND RIGHT OCULAR REGION WHEN THE PRE-TRAINED MODELS ARE FUSED USING FEATURE LEVEL FUSION

Model	Enrollment	Verification	EER for left ocular region(%)	EER for right ocular region(%)
Feature fusion of VGG16, ResNet18, ResNet-like model	Office	Office	18.32	18.59
	Office	Daylight	33.15	28.10
	Office	Dim	38.61	36.81
	Daylight	Daylight	19.49	17.97
	Dim	Dim	13.03	11.23

TABLE VII
EERS FOR FACE WHEN THE PRE-TRAINED MODELS ARE FUSED USING FEATURE LEVEL FUSION

Model	Enrollment	Verification	EER for face region(%)
Feature fusion of VGG16, ResNet18, MobileNet, LightCNN9	Office	Office	18.88
	Office	Daylight	28.30
	Office	Dim	31.83
	Daylight	Daylight	18.38
	Dim	Dim	14.83

Table VI and Table VII show that EERs decreased on average for both left and right ocular regions and face regions, when VGG16, ResNet18, and ResNet18-like model were fused using feature level fusion for ocular region, and VGG16, ResNet18, MobileNet V1, MobileNet V2 and LightCNN9 were fused using feature level fusion for face region. On an average, the EERs were decreased.

IV. CONCLUSION AND FUTURE WORK

The main aim of this work was to evaluate hand-crafted and deep features from pre-trained models across illumination conditions for fusion of selfie face and ocular regions. As a part of future work, these pre-trained models may be fine-tuned and other learning based models may be used to decrease the EER rates. Reducing the computational efficacy footprint is another improvement avenue.

REFERENCES

[1] Rattani, A. and Derakhshani, R. (2017). Ocular biometrics in the Visible spectrum: A survey. Image and Vision Computing, 59, pp.1-16.

[2] Jain, A., Nandakumar, K. and Ross, A. (2005). Score normalization in multimodal biometric systems. Pattern Recognition, 38(12), pp.2270-2285.

[3] A. Rattani and R. Derakhshani, ”Online co-training in mobile ocular biometric recognition,” 2017 IEEE International Symposium on Technologies for Homeland Security (HST), Waltham, MA, 2017, pp. 1-5.

[4] K.He, X.Zhang, S. Ren, and J.Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770 778.

[5] Alpaydin, E. (2014). Introduction to machine learning. Cambridge (USA): MIT Press.

[6] A. Rattani, R. Derakhshani, ”On fine-tuning convolutional neural networks for smartphone based ocular recognition”, IEEE International Joint Conference on Biometrics (IJCB), pp. 762-767, Oct 2017.

[7] A. Rattani, R. Derakhshani, ”A survey of mobile face biometrics”, Computers and Electrical Engineering, vol. 72, pp. 39-52, 2018.

[8] A. Rattani, R. Derakhshani, S. K. Saripalle, V. Gottemukkula, ”ICIP 2016 competition on mobile ocular biometric recognition”, IEEE Intl Conf. on Image Processing (ICIP) 2016 Challenge Session on Mobile Ocular Biometric Recognition, pp. 320-324, 2016.

[9] A. Rattani, D. R. Kisku, M. Bicego, M. Tistarelli, ”Robust feature-level multibiometric classification”, 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference, pp. 1-6, Sept 2006.

[10] A. Rattani, N. Reddy, R. Derakhshani, ”Convolutional neural network for age classification from smart-phone based ocular images”, IEEE International Joint Conference on Biometrics, pp. 756-761, Oct 2017.

[11] Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[12] Rattani, Ajita Derakhshani, Reza Gottemukkula, Vikas. (2016). ICIP 2016 competition on mobile ocular biometric recognition. 320-324. 10.1109/ICIP.2016.7532371.