Attribute Estimation Using Multi-CNNs from Hand Images

Yi-Chun Lin*, Yusei Suzuki†, Hiroya Kawai†, Koichi Ito†, Hwann-Tzong Chen*, and Takafumi Aoki†
* Department of Computer Science, National Tsing-Hua University, Taiwan.
† Graduate School of Information Sciences, Tohoku University, Japan.
E-mail: ito@aoki.ecei.tohoku.ac.jp

Abstract—**The human hand is one of the primary biometric traits in person authentication. A hand image also includes a lot of attribute information such as gender, age, skin color, accessory, and etc. Most conventional methods for hand-based biometric recognition rely on one distinctive attribute like palmprint and fingerprint. The other attributes as gender, age, skin color and accessory known as soft biometrics are expected to help identify individuals but are rarely used for identification. This paper proposes an attribute estimation method using multi-convolutional neural network (CNN) from hand images. We specially design new multi-CNN architectures dedicated to estimating multiple attributes from hand images. We train and test our models using 11K Hands, which consists of more than 10,000 images with 7 attributes and ID. The experimental results demonstrate that the proposed method exhibits the efficient performance on attribute estimation.**

I. Introduction

Biometric recognition is a technique of identifying individuals using their behavioral or physiological characteristics, for example, fingerprint, face, iris, palmprint, voice, gait, signature, etc. [1]. Some of biometric recognition systems using fingerprint, face, iris and palmpirnt have been available in person authentication. Such systems still have a problem under practical situations, since their performance decreases due to pose changes, illumination changes, low-resolution images, human motion, etc. One of performance improvement approaches for biometric recognition is to use attribute information, where biometric recognition using attributes is called soft biometrics [2], [3].

An image of biometric traits, e.g., a face image, includes a lot of attribute information such as age, gender, ethnicity, hair color, nose size, mouth shape, etc., which can be used as low- or mid-level features in face recognition [4], [5], [6]. There are only a few methods of attribute estimation for biometric recognition except for face recognition, although soft biometrics is one of the important research topics in biometric recognition for practical use. In this paper, we focus on palmprint recognition [7], [8], since a palm image can be taken under contactless situation and the recognition performance is comparable with that of fingerprint recognition.

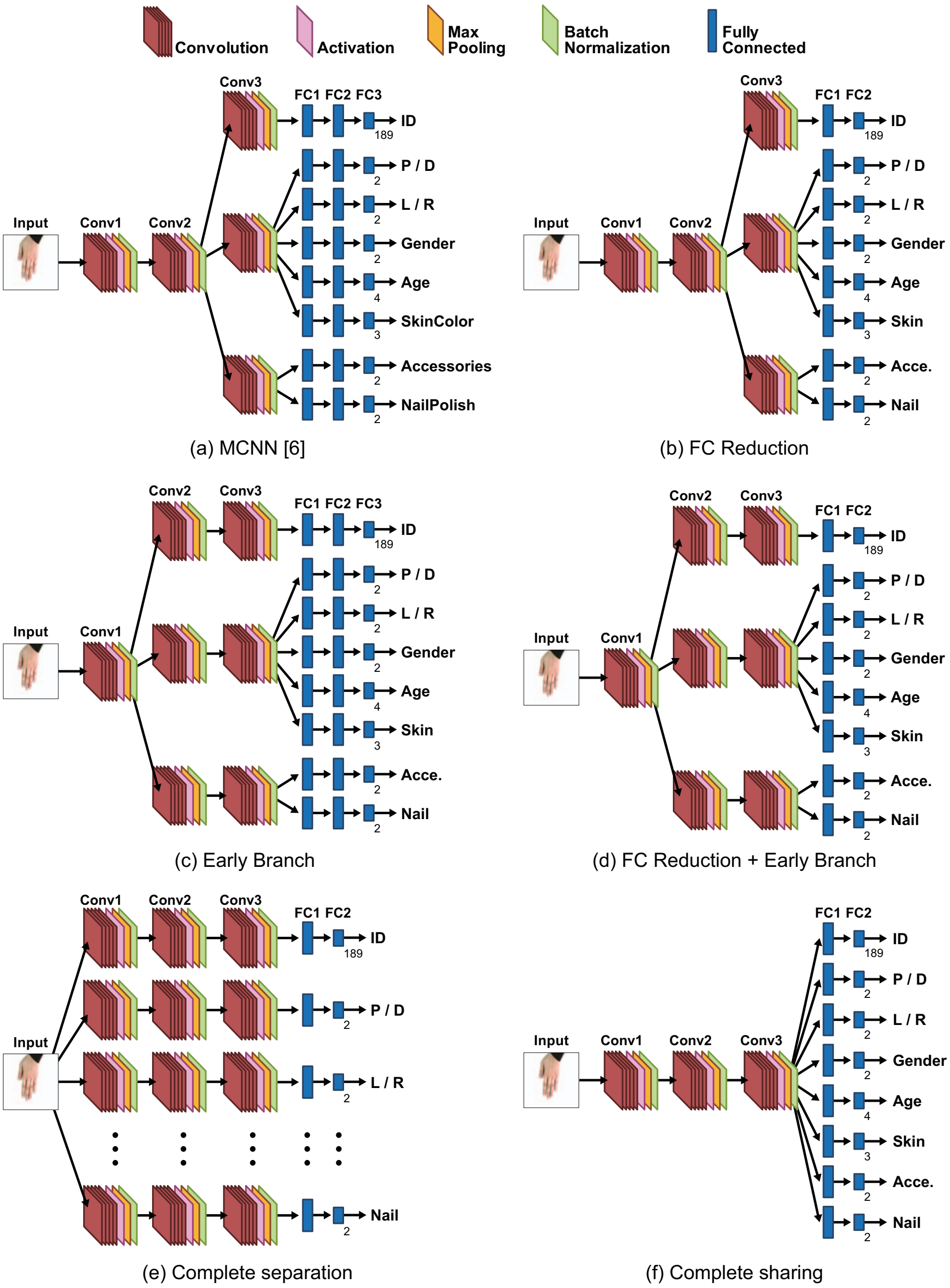
A hand image having five fingers and a palm region (front and back surface) also includes some distinctive attribute information such as age, gender, skin color, etc. There are some attribute estimation methods from hand images [9], [10], [11], [12], [13], which focus only on gender estimation and consider forensic applications. In addition, it is difficult to manually design adequate features for estimating attributes from hand images due to their wide variety of features. Hence, a deep learning-based approach is suitable in attribute estimation as well as face recognition [4], [5], [6]. Afifi [14] proposed a method using convolutional neural network (CNN) in substitution for traditional feature extractors. This method employs a two-stream CNN, whose architecture is based on AlexNet [15]. The input for the first CNN is a blurred image after applying a bilateral filter and that for the second CNN is an edge image after applying a guided filter. The CNN model is trained for gender recognition. Features extracted from CNN is input to SVM classifiers and the final score is obtained by the sum of classifiers. This method focuses on person authentication and the CNN model is not always suitable in attribute estimation. There is no literature of focusing on attribute estimation from hand images to the best of our knowledge.

In this paper, we propose an attribute estimation method using multi-CNNs from hand images. We specially design new multi-CNN architectures dedicated to estimating multiple attributes from hand images, which are inspired by a CNN architecture for multi-task learning [6] used in face attribute estimation. We considered 6 multi-CNN architecture, where we adopted the main concept of MCNN architecture and perform some modifications for hand images. We use 11k Hands[1], which is a large hand image dataset with attributes. Through a set of experiments using 11k Hands, we demonstrate that the proposed method exhibits good performance on attribute estimation compared with other methods.

II. Proposed Method

Face attribute estimation has achieved great success by using CNN with multi-task learning. Hand et. al. [6] proposed a face attribute estimation method using a multi-task convolutional neural network (MCNN), whose architecture is based on the correlation among face attributes. It is expected that this approach of using attribute correlation is also effective in hand attribute estimation. Hence, we consider new CNN architectures dedicated to estimating multiple attributes from hand images inspired by MCNN. We considered 6 multi-CNN architecture as shown in Fig. 1, where we adopted

[1] https://sites.google.com/view/11khands



(a) MCNN [6]

(b) FC Reduction

(c) Early Branch

(d) FC Reduction + Early Branch

(e) Complete separation

(f) Complete sharing

Fig. 1. Multi-CNN architectures proposed in this paper.

the main concept of MCNN architecture and perform some modifications for hand images.

The first modification is according to the relation among attributes. We present a simple model assuming that attributes are highly related each other. In this case, we can use the architecture containing only one feature extractor for all the attributes as shown in Fig. 1 (f). This architecture is referred to "Complete sharing". For comparison, another base model is designed with a completely parallel architecture called "Complete separation", where each feature extractor learns only one attribute information as shown in Fig. 1 (e).

The second modification is attribute grouping. Unlike facial attributes, hand attributes share regional information. A hand image can be separated into fingers and a palm or dorsal region to extract distinctive area features. Attributes such as gender and age belong to a palm or dorsal region, while other attributes such as necessaries and nail polish belong to finger regions. Hence, we can make two major groups for region feature extractor. Note that ID is not included in both groups, since the ID feature contains more personal information, which are often regarded as bias comparing to common features. Therefore, we treat ID as the third group. This architecture is designed based on MCNN as shown in Fig. 1 (a) and is called "MCNN". We also consider removing one FC layer from "MCNN", since the size of hand datasets is much smaller than that of face datasets. The redundant parameters may lead to over-fitting of the model, and reducing the number of parameters is for compact implementation. This architecture is shown in Fig. 1 (b) and is called "FC Reduction".

The last modification is the amount of shared information. For both "FC Reduction" and "MCNN", the feature extractors concentrate more on information common to all the attributes rather than representational information of each attribute, since there are two shared layers. Considering that the unique information may provide more assistance on each attribute estimation, we intensify the learning of region features by reducing one shared layer for both "FC Reduction" and "MCNN". The two new models are called "Early Branch" and "FC Reduction + Early Branch" and are shown in Fig. 1 (c) and (d), respectively.

For the parameters of the models, Conv1 possesses 75 7x7 convolution filters, Conv2 possesses 200 5x5 filters and Conv3 possesses 300 3x3 filters. The stride of all convolutional layers is 1. Each convolutional layer is followed by a ReLU, 3x3 Max Pooling with stride 3 and Batch Normalization. Except for the last FC layers, all the FCs have 512 units and are followed by a ReLU with a 30% dropout. For the last FCs, the units are depended on the class of each attribute.

III. Experiments

A. Dataset

There are some hand image datasets with labeled attributes such as Sun et al. [16], Yoruk et al. [17], Hu et al. [18], Kumar [19], Ferrer et al. [20], and Afifi [14]. Most datasets provide only a few attributes and insufficient number of hand images for CNNs. Among them, 11k Hands[1] provided by Afifi [14] consists of more than 10,000 images with 8 attributes. Therefore, we used 11K Hands in the following experiments. Fig. 2 shows the samples of each attribute. This dataset consists of 189 subjects and 58 images per person. The total number of hand images is 11,076 and the image size is $1,600 \times 1,200$ pixels. 11K Hands contains 8 attributes: (i) the subject ID, (ii) gender, (iii) age, (iv) skin color, (v) hand side (dorsal and palmar), (vi) right or left hand, (vii) accessories, and (viii) nail polish. The subjects range in age from 18 to 75. In our experiments, the age is divided into 4 classes: $18 \sim 30$, $31 \sim 45$, $46 \sim 60$, and $61 \sim 75$. The skin color is divided into 4 groups in the original dataset, which are "very fair", "fair", "medium", and "dark". In our experiments, we integrate "very fair" into "fair" to avoid the bias, since the number of "very fair" subjects is only 5.

B. Experimental settings

The experiments are implemented on an Intel Xeon CPU E5 - 1650 v4 @ 3.60 Hz CPU, 64GB RAM and NVIDIA GTX 1080 Ti GPU. We compare the performance on attribute estimation with all the CNN architectures using the 11k Hands dataset. All the methods are implemented using PyTorch.

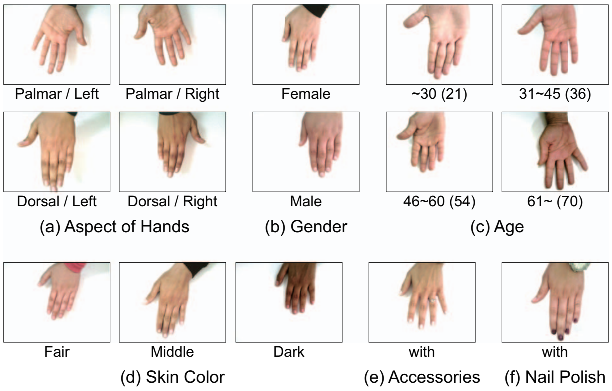


Fig. 2. Example of hand images for each attribute in 11k Hands.

We implement data augmentation on the dataset including random crop and random rotation with degree range between plus and minus 10 degrees. Then, the hand images are resized to 224×224 pixels and their pixel values are normalized in terms of mean and variance, which are optimized in ImageNet data. The loss is calculated by cross entropy and the initial learning rate is 0.001, which is controlled by the PyTorch function "CosineAnnealingLR", where the maximum number of iterations is 100. We use Nesterov Accelerated Gradients Optimizer [21]. We train the network for both attribute estimation and identification for maximum 100 epochs and use a batch size of 8. To avoid overfitting, early stopping is activated when the loss on the validation data does not improve in 10 consecutive epochs.

We divide the dataset into dorsal set and palmar set. In each set, we subsample 4 training images and 1 validation image per person from 189 subjects. The total number of used images is 1,512 for training and 378 for validation. From the remaining images in each set, we randomly select 2 images per person for testing. The process of evaluation is repeated 10 times for all the attributes and calculate the average accuracy.

C. Experimental results

Table I summarizes the experimental results. From the results of "Complete sharing", we validate attribute features share some common features so that weight sharing in MCNN can be applied to estimating multiple attributes. However, comparing with the other models, one feature extractor is insufficient to cover all kinds of attribute information. This fact indicates that dedicated feature extractors for attributes are required. "Complete separation" applied completely independent feature extractors for each attribute shows a slight improvement in most attributes and an unexpected increasing of ID accuracy. This result informs that ID possesses more unique information and shares fewer features with the other attributes. To recognize ID from hand images, ID should be separated from the other attributes if weight sharing is applied. From the results of the two models, it implies that the proper utilization of weight sharing is able to replace some of the feature extractors without sacrificing the performance.

TABLE I
ACCURACY [%] OF ATTRIBUTE ESTIMATION AND IDENTIFICATION FOR THE 11K HANDS DATASET

Method	P/D	L/R	Gender	Age	SkinColor	Accessories	NailPolish	Ave. Att.	ID
MCNN [6]	99.47	99.42	94.07	98.99	89.26	99.52	97.28	96.86	74.81
FC Reduction	99.52	99.60	93.65	99.10	87.72	99.50	96.01	96.44	75.71
Early Branch	99.87	**99.76**	94.15	**99.42**	89.39	**99.63**	97.49	97.10	75.42
FCR + EB	99.71	99.63	93.10	99.37	**90.21**	99.55	**98.23**	**97.11**	**78.97**
Complete separation	**99.89**	99.66	94.07	99.34	85.77	99.50	97.86	96.58	75.40
Complete sharing	99.18	99.02	**94.31**	99.13	84.71	99.23	95.50	95.87	71.90

We perform weight sharing by grouping attributes based on regions and adding a branch for ID in the remaining four models. The results from the models prove the possibility of combining weight sharing with ID separation. We further focus on the influence of shared information and FC layers. While "MCNN" and "FC Reduction" contain two shared layers, there is only one shared layer for "Early Branch" and "FCR + EB". One shared layer has better performance than two shared layers as observed in the results. It is reasonable that regional features include detailed information for each attribute rather than hand features. By focusing more on each attribute information, models can estimate attributes more precisely. Since the scale of weight sharing is confirmed, we consider the number of FC layers. According to the original MCNN, there might exist overfitting in terms of redundant parameters, which result in lower accuracy. Since we trained our models on a rather small dataset, "FC Reduction" and "FCR + EB" are evaluated to verify the potential risk of overfitting. As demonstrated, "FCR + EB" achieves the highest accuracy of ID estimation and the highest average accuracy of attributes and both models have small performance improvement of other attributes comapred with the models having two FC layers. This result indicates that the overfitting occurs in all the attributes and has a substantial impact on ID. Although the overfitting is prevented from removing one FC layer, the accuracy of ID does not exceed 80%. We conclude that ID estimation is a difficult problem, since ID estimation requires a special architecture to extract more representational information.

IV. CONCLUSION

We proposed an attribute estimation method using multi-CNNs from hand images. We designed new CNN architectures dedicated to estimating multiple attributes from hand images. Through a set of experiments using 11k Hands, we demonstrated that the proposed method of "FCR + EB" exhibits the good performance on attribute estimation compared with other multi-CNN architectures.

REFERENCES

[1] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*, Springer, 2008.

[2] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," *Proc. Int'l Conf. BIometric Authentication (LNCS 3072)*, pp. 731–738, 2004.

[3] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Information Forensics and Security*, vol. 11, no. 3, pp. 441–467, Mar. 2016.

[4] Y. Zhong, J. Sullivan, and H. Li, "Face attribute prediction using off-the-shelf CNN features," *Proc. Int'l Conf. Biometrics*, June 2016.

[5] M. Ehrlich, T. J. Shields, T. Almaev, and M. R. Amer, "Facial attributes classification using multi-task representation learning," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops*, pp. 47–55, June 2016.

[6] E. M. Hand and R. Chellappa, "Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification," *Proc. the Thirty-First AAAI Conf. Artificial Intelligence*, pp. 4068–4074, Feb. 2017.

[7] D. Zhang, *Palmprint Authentication*, Kluwer Academic Publication, 2004.

[8] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *Pattern Recognition*, vol. 42, no. 7, pp. 1408–1418, Jan. 2009.

[9] J. L. Scheuer and N. M. Elkington, "Sex determination from metacarpals and the first proximal phalanx," *J. Forensic Sci.*, vol. 38, no. 4, pp. 769–778, 1993.

[10] R. A. Lazenby, "Identification of sex from metacarpals:Effect of side asymmetry," *J. Forensic Sci.*, vol. 39, no. 5, pp. 1188–1194, 1994.

[11] A. B. Falsetti, "Sex assessment from metacarpals of the human hand," *J. Forensic Sci.*, vol. 40, no. 5, pp. 774–776, 1995.

[12] G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classification from hand shape," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*, pp. 1–7, June 2008.

[13] T. Kanchan and K. Krishan, "Anthropometry of hand in sex determination of dismembered remains - A review of literature," *J. Forensic and Legal Medicine*, vol. 18, no. 1, pp. 14–17, Jan. 2011.

[14] M. Afifi, "Gender recognition and biometric identification using a large dataset of hand images," *CoRR*, vol. abs/1711.04322, pp. 1–19, 2017.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proc. Annual Conf. Neural Information Processing Systems*, pp. 1–9, 2012.

[16] Z. Sun, T. Tan, Y. Wang, and S. Z. Li, "Ordinal palmprint representation for personal identification," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 279–284, June 2005.

[17] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Trans. Image Processing*, vol. 15, no. 7, pp. 1803–1815, July 2006.

[18] R.-X. Hu, W. Jia, D. Zhang, J. Gui, and Song. L.-T., "Hand shape recognition based on coherent distance shape contexts," *Pattern Recognition*, vol. 45, no. 9, pp. 3348–3359, Sept. 2012.

[19] A. Kumar, "Incorporating cohort information for reliable palmprint authentication," *Proc. 2008 Sixth Indian Conf. Computer Vision, Graphics & Image Processing*, pp. 583–590, Dec. 2008.

[20] M. A. Ferrer, A. Morales, C. M. Travieso, and J. B. Alonso, "Low cost multimodal biometric identification system based on hand geometry, palm and finger print texture," *Proc. 2007 41st Annual IEEE Int'l Carnahan Conf. Security Technology*, pp. 52–58, Oct. 2007.

[21] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O(1/k2)," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.