Hybrid Feature Learning and Engineering Based Approach for Face Shape Classification

Theiab Alzahrani
Department of Electrical Engineering and Electronics
University of Liverpool
Liverpool, UK
pstalzah@liverpool.ac.uk

Waleed Al-Nuaimy
Department of Electrical Engineering and Electronics
University of Liverpool
Liverpool, UK
wax@liverpool.ac.uk

Baidaa Al-Bander
Department of Computer Engineering
University of Diyala
Diyala, Iraq
baidaa.q@gmail.com

***Abstract*—Face shape classification is a vital process to choose an appropriate eyelashes, hairstyle and facial makeup, and selection of a suitable glasses' frames according to the guidelines from experts. Measuring face characteristics by beauty experts manually costs time and efforts. Therefore, developing automated face shape identification system could alleviate the need for additional time and efforts made by experts. Many automatic face shape classification methods have been proposed in the literature; however, the existing methods tackle many challenges due to the complexity of face geometry and variation in its characteristics. This paper presents a deep convolutional neural network method (CNN) approach for classifying face shape into five types. The proposed method which is based on merging the features learnt by CNN with hand crafted features represented by histogram of oriented gradients (HOG) and facial landmarks has proven to be efficient in identification of facial shape. The obtained results demonstrate that the proposed method is promising in identifying the shape of face achieving accuracy of 81.1%.**

***Index Terms*—Face shape, Convolutional neural networks, Cosmetics, Hand-crafted features, Engineered features**

I. Introduction

Identifying of person's face shape plays important role in recommending the suitable hairstyle, eyelashes, makeup, and glasses by fashion stylists. The process of determining facial shape manually by beauty experts can be carried out in several stages such as capturing pictures, outlining the face, measuring the width and length of the face, jaw, forehead, cheekbones, and finally determining of the face shape. According to the beauty experts, the face shape can be recognised as five categories namely: oval, square, round, oblong, and heart shapes [1].

With a noticeable growth of computer aided design systems, automation the process of face classification based on image processing and computer vision strategies can help in decreasing time and efforts achieved by experts. In the literature, many automated face shape classification systems were presented. Many of these published face classification methods consider extracting the face features manually then passing them to three classifiers for classification including Discriminant Analysis (LDA), Artificial Neural Networks (ANN), and Support Vector Machines (SVM) [2], SVM [3], k-Nearest Neighbours [4], and Probabilistic Neural Networks [5]. Furthermore, Bansode et al. [6] proposed face shape identification method based on three criteria which are region similarity, correlation coefficient and fractal dimensions. Recently, Pasupa et al. [7] presented a hybrid approach combined VGG convolutional neural network (CNN) with SVM for face shape classification. Moreover, Emmanuel [8] adopted pretrained Inception CNN for classifying face shape using features extracted automatically by CNN.

The work presented by researchers showed progress in face shape interpretation, however, existing face classification systems require more effort to achieve a better performance. Face shape identification is a challenging duty due to the complexity of face and the possible variations in rotation, size, illumination, hairstyle, age and expressions. Furthermore, the existence of face occlusion from hats and glasses also adds difficulties to the classification process.

In this work, we propose deep learning-based approach for automated face shape classification into heart-shaped, oblong, oval, round, and square. The developed method combines the features engineered manually such as histogram of oriented gradients (HOG) and facial landmarks with the features extracted automatically by convolutional neural network to identify the face shape. The remaining of this paper is presented as follows: in Section II, the materials and proposed framework are described and explained; results of the proposed system are revealed and discussed in Section III, and finally, the work has been concluded in Section IV.

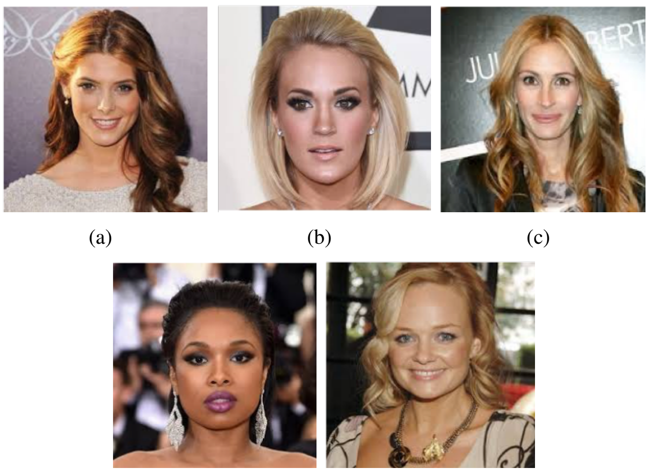
T. Alzahrani is financially supported by the Kingdom of Saudi Arabia government.

978-1-7281-4813-7/19/$31.00 ©2019 IEEE

II. Materials and Method

A. Materials

The dataset collected and made available on request by the author in [8] are used to evaluate the proposed system. This database comprises of 500 female celebrity images which has been labelled by an expert into five face shapes namely heart-shaped, oval, oblong, square and round. Fig. 1 shows samples of face shape types.



(a) (b) (c) (d) (e)

Fig. 1. Sample images showing the five face shapes. (a) Heart-shaped face (b) Oblong-shaped face (c) Oval-shaped face (d) Round-shape face (d) Square-shape face

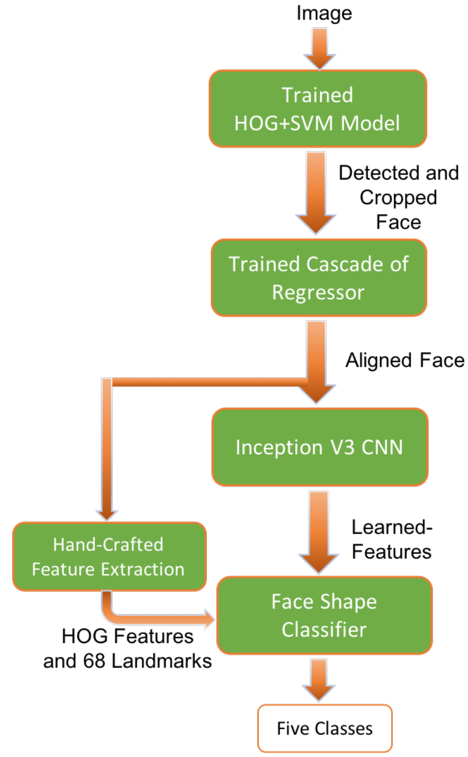
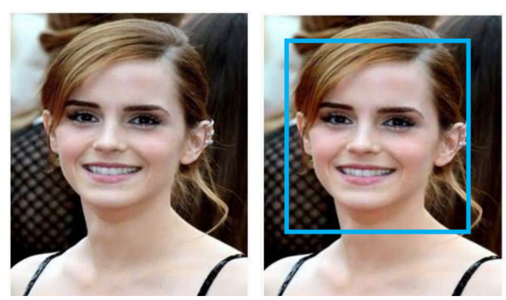


Fig. 2. Block diagram of proposed face shape classification system

B. Method

The block diagram of proposed system shown in Fig. 2 including three main phases are explained as follows:

1) **Face Detection and Cropping**: In this stage of face shape classification framework, face is detected and cropped from an image as shown in Fig. 3. The face detection is carried out using a widely applied face detection model trained on histogram of oriented gradients (HOG) features with a linear classifier; Support Vector Machine (SVM) model. HOG based method [9] is typically more accurate than Haar cascades [10] with less false positives and requires less parameters to tune at test time. The dataset used for training, consists of 2825 images which are obtained from LFW dataset [11].
HOG algorithm iterates on every pixel of an input image, in each pixel, it checks all the pixels around it and figures out how dark the current pixel is compared with the adjacent pixels that surrounding it. It then measures in which direction the pixel is getting darker and this direction can be called gradient. This procedure is repeated on every single pixel. The image is then broken-down into 16×16 squares, where number of gradients point in each major direction is counted. That square is typically replaced with the direction which was the strongest. Hence, the image is converted into representation that captures the basic structure of a face. These features represented by face representations are used to train a Linear SVM classifier to detect region of faces in a test image.



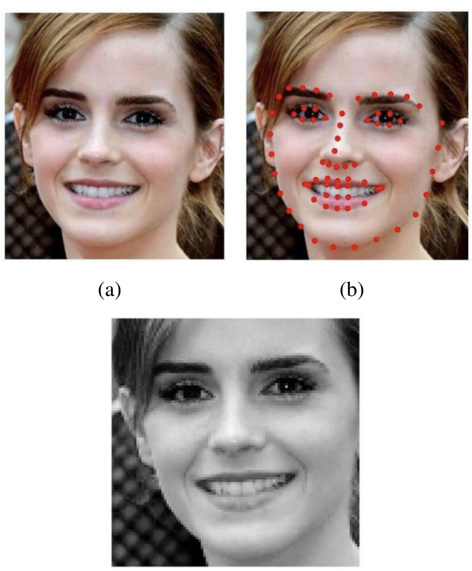
(a) (b)

Fig. 3. (a) Original image (b)Face region localised to be cropped.

2) **Landmark Detection and Face Alignment**: In this stage, the detected and cropped face is aligned by firstly detecting the face landmarks (68 landmarks) by ensemble of regression tree method (ERT) [12] and then aligning the face using the detected landmarks as shown in Fig. 4. Face alignment is a form of data normalisation. The face alignment is similar to feature vectors normalisation done via scaling to unit norm or zero centring prior to learning a machine learning algorithm. It is common to align the faces in dataset before training a face recogniser.
The face landmark detection algorithm; subsequently

face alignment, adopted in this work is an implementation of the ensemble of regression trees (ERT) presented in 2014 by Kazemi and Sullivan [12]. To directly estimate the landmark positions, this method uses simple and fast feature which is pixel intensities differences. Cascade of regressors are adapted to refine these estimated positions with an iterative process. At each iteration, the regressors produces a new estimate from the previous one, trying to reduce the alignment error of the estimated points. To obtain a normalised translation, rotation, and scale representation of the face, the alignment procedure relies on the facial landmarks; in particular the eye regions.

The detected facial landmarks coordinates (68 landmarks) help to align the faces by making all faces in the dataset centred in the image, scaled such that the size of the faces is almost identical, and rotated such that the eyes locate on a horizontal line (the face is rotated such that the eyes locate along the same y-axis). Ensemble of regression trees (ERT) [12] for landmark detection is trained on 300W dataset [13] which contains more than 4000 images in the wild.



(a) (b)

(c)

Fig. 4. Emma Watson's face is correctly identified as an oval face shape. (a) Cropped face (b) Detected 68 landmarks (c) Aligned face.

3) **Classification by Inception CNN**: In the final stage, the aligned images are used for training and evaluating the model. The training images are fed to the Inception V3 convolutional neural network [14] along with HOG features and landmarks to classify face shape into five classes as shown in Fig. 5 .

GoogleNet (Inception v3) CNN architecture [14] is adopted and adapted to identify the shape of face. The transfer learning strategy which is proven to be effective in several cases is considered rather than training net from scratch. The Inception model, which is pre-

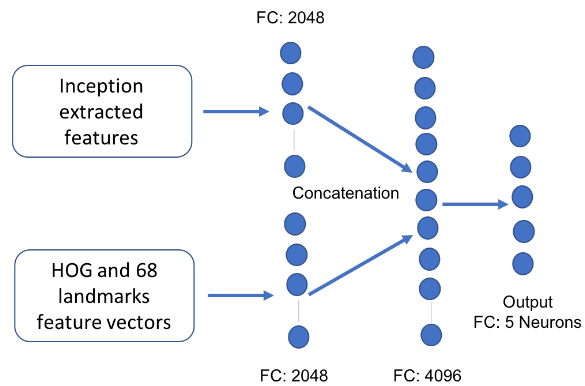


Fig. 5. Block diagram shows merging Inception with handcrafted features in the face shape classifier.

trained on ImageNet [15], is exploited by considering the pre-trained weights and then fine-tuning them. As Inception was originally trained on 1000 class (i.e 1000 output neurons), the last layer (fully connected layer) is truncated and replaced with a fully connected layer with 5 output neurons (5 types of face shape).

The Inception model comprises five convolutional layers alternated with max-pooling operations, successive stacks of 11 Inception modules which are basically mini-models inside the bigger model, and softmax ouput layer and auxiliary softmax as an intermediate output. Stacking several Inception modules makes the network architecture complicated and deep. In addition to the depth of the network, Inception modules are also wide which are designed to detect features at multiple length scales. The size of convolutional filters in each block are either 1×1, 3×3, 5×5 , 1×7 or 7×1.

Initially, 80% of the data are randomly chosen for training and the rest 20% were used for independent testing. The pre-trained Inception network are fine-tuned with 200 epochs on images resized into 299×299 pixels. To train the network, the stochastic gradient descent SGD with the momentum algorithm was adopted for optimisation with learning rate of 0.016 and momentum parameter 0.95.

The HOG and 68 face landmarks (hand-designed features) - extracted from the same training images which are used to train Inception CNN - are passed into fully connected layer (FC) comprises 2048 neurons. The output of FC layer in Inception which has 2048 nodes are concatenated with FC layer of handcrafted features producing layer of 4096 neurons. This resulted fully connected layer is followed by fully connected layer of 5 neurons and softmax layer for classification purpose.

III. Results and Discussion

The experiments have been carried out on an HP Z440 with Linux Mint OS with RAM of 16GB , an Intel Xeon E5 3.50 GHz processor and NVIDIA GTX TITAN X GPU card with 12GB and 3072 cores of CUDA parallel-processing. The obtained results from developed system are shown in Table I. From the table, it can be noticed that face shape classification using features extracted by CNN, CNN features combined with HOG features, CNN features combined with face landmarks, CNN features combined with HOG and face landmarks achieve accuracy of 75.2%, 76.9%, 78.2%, and 81.1%, respectively.

The results reveal that combining hand crafted features with automatically extracted features can improve the overall performance of the face shape classification system. It also shows that our approach outperforms the other methods in the literature. However, due to the lack of datasets and their ground truth used in the existing methods (private dataset), it was not possible to test our proposed framework on other datasets. Author in [8], who proposed system based on passing the images directly into Inception CNN for feature extraction and classification, reported accuracy of 84.4% on the same data that we used for training and testing the proposed method. However, the author tested his method on the whole data including data used for training which results in overfitting. Furthermore, testing on the data used for training is not an indicator for the system performance. In our system, 80% of data was used for training while the remaining 20% of unseen data was retained for testing.

TABLE I

THE PERFORMANCE OF PROPOSED FACE SHAPE CLASSIFICATION METHODOLOGY COMPARING TO THE EXISTING METHODS IN LITERATURE.

Method	Accuracy	Dataset
Inception V3 [8]	84.4%	500 image [8]
Region Similarity, Correlation and Fractal Dimensions [6]	80%	Caltech 450 [16], Cuhk 260 [17], Lfw 200 [11]
Active Appearance Model (AAM), segmentation, and SVM [2]	72%	1000 image [2]
Hybrid approach VGG and SVM [7]	70.3%	500 image [7]
3D face data and SVM [3]	73.68%	209 3D image [3]
Geometric features [4]	80%	400 image [4]
Probability Neural Network and Invariant Moments [5]	80%	120 image [5]
Inception3 CNN	75.2%	500 image [8]
Inception3 CNN+HOG	76.9%	
Inception3 CNN+LMs	78.2%	
Inception3 CNN+HOG+LMs	**81.1%**	

Finally, it is worth mentioning that the results obtained from our proposed system require more improvements. This is due to the limited and small number of training and testing samples used to train and test the model which are only 500 labelled images. It also needs more verification for the generalisation capabilities of the proposed system when it comes to test images outside this dataset. Therefore, we aim to increase the labelled data which is currently part of the authors' ongoing and current research.

IV. CONCLUSIONS

In this work, a framework for face shape classification has been presented and evaluated. The proposed method has proven that merging the features extracted manually with features learned by CNN can boost the classification performance. The results obtained from the proposed model comparing to the existing method proved the efficiency of presented system for face shape classification. This work can be extended, and results can be improved by providing more labelled data for study as well as testing more handcrafted features and more complicated CNN architectures.

REFERENCES

[1] J. Derrick, "Is your face round, square, long, heart or oval shaped?" https://www.byrdie.com/is-your-face-round-square-long-heart-or-oval-shaped-345761, accessed: 2019-06-01.

[2] W. Sunhem and K. Pasupa, "An approach to face shape classification for hairstyle recommendation," in *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, 2016, pp. 390–394.

[3] P. Sarakon, T. Charoenpong, and S. Charoensiriwath, "Face shape classification from 3D human data by using SVM," in *The 7th 2014 Biomedical Engineering International Conference*. IEEE, 2014, pp. 1–5.

[4] A. Zafar and T. Popa, "Face and eye-ware classification using geometric features for a data-driven eye-ware recommendation system," in *Graphics Interface Conference*. ACM Library, 2016, pp. 183–188.

[5] R. F. Rahmat, M. D. Syahputra, U. Andayani, and T. Z. Lini, "Probabilistic neural network and invariant moments for men face shape classification," in *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 1. IOP Publishing, 2018, p. 012095.

[6] N. Bansode and P. Sinha, "Face shape classification based on region similarity, correlation and fractal dimensions," *International Journal of Computer Science Issues (IJCSI)*, vol. 13, no. 1, p. 24, 2016.

[7] K. Pasupa, W. Sunhem, and C. K. Loo, "A hybrid approach to building face shape classifier for hairstyle recommender system," *Expert Systems with Applications*, vol. 120, pp. 14–32, 2019.

[8] E. D. Tio., "Face shape classification using inception v3," https://github.com/adonistio/inception-face-shape-classifier, 2017, accessed: 2019-06-01.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition (CVPR'05)*, vol. 1. IEEE Computer Society, 2005, pp. 886–893.

[10] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[11] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[12] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

[13] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[16] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1513–1520.

[17] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2008.