STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LXVI**, Number 2, 2021
DOI: 10.24193/subbi.2021.2.05

AUTOMATIC FACE SHAPE CLASSIFICATION VIA FACIAL LANDMARK MEASUREMENTS

ALEXANDRU-ION MARINESCU

Abstract. This paper tackles the sensitive subject of face shape identification via near neutral-pose 2D images of human subjects. The possibility of extending to 3D facial models is also proposed, and would alleviate the need for the neutral stance. Accurate face shape classification serves as a vital building block of any hairstyle and eye-wear recommender system. Our approach is based on extracting relevant facial landmark measurements and passing them through a naive Bayes classifier unit in order to yield the final decision. The literature on this subject is particularly scarce owing to the very subjective nature of human face shape classification. We wish to contribute a robust and automatic system that performs this task and highlight future development directions on this matter.

1. Introduction

Of the major areas of application of the topic of face shape classification, we will mention the most prominent two: hairstyle or eye-wear *recommender systems* and forensic analysis of human subjects, by complementing *3D facial reconstruction*.

Recommender systems are first and foremost an important marketing tool and a major revenue source for the fashion and entertainment business sectors. They seek, aided through computing processing power, to mimic the way the potential customer thinks, by keeping track of the products she/he finds interesting. To put it simply, they create a psychological profile of the customer, attuned for the target product category. There exist a plethora of recommender system types: some are trained for music or movie recommendation, based on music genre (i.e. classical, pop, jazz, rock) or movie category

Received by the editors: 15 September 2021.

2010 *Mathematics Subject Classification.* 68T45.

1998 *CR Categories and Descriptors.* I.2.10 Computing Methodologies [**ARTIFICIAL INTELLIGENCE**]: Vision and Scene Understanding – *Modeling and recovery of physical attributes* .

Key words and phrases. artificial intelligence, computer vision, naive Bayes, face shape, facial landmarks.

(i.e. horror, drama, comedy, action) preferences, whilst others such as the ones employed by major online stores, attempt to track what the end client would be interested in buying next. Another possible application of a face shape classifier would be in the field of forensic analysis. Here, for example, a suspect's face shape could serve as a hash check for fast querying against a police database of known criminals.

Nevertheless, our particular focus in this paper will be on a hairstyle and eye-wear recommender system. More specifically, we will discuss the implementation of face shape classifiers, which serve as the basic building block of such an application. Face shape recognition has become very useful in many computer vision applications. So, an algorithm to classify the face shape correctly is needed. There can be issues if the images are not of good quality and have pose variability. We aim to distinguish seven types of face shapes: oval, round, rectangle, square, heart, diamond and triangular (see Figure 1). The face shape is to be analyzed from the frontal/neutral pose. Cancelling the yaw, pitch and roll of the subject's face has been discussed previously in [3, 7]. In the following sections we will describe how this can be accurately achieved using a combination of facial landmark measurements in the standard 68-landmark model (Figure 2) and train a naive Bayes classifier in order to yield the final decision regarding the user's face shape.

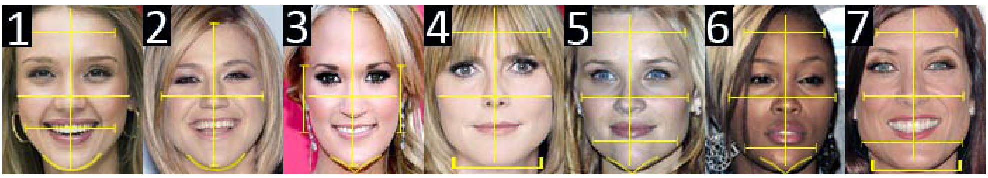

Figure 1. The 7 generally acknowledged face shapes, in reading order from left to right: oval, round, rectangle, square, heart, diamond and triangle (*thehairstyler.com*).

2. State of the art

The subject of face shape classification is a difficult one mainly due to the fact that determining the face shape of a human is very subjective and open to interpretation. In general, a person does not belong strictly to one of the seven classes of shapes, but instead, possesses a combination of at least two principal shapes. At most, what we can say is that a person has "*predominantly*" the facial traits of a certain category. As a consequence, a standardized face shape classification is yet to be developed. Some sources suggest fewer face

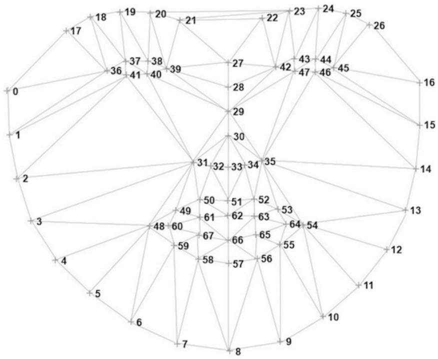

FIGURE 2. The 68 landmark-based face model, which serves as input for our face shape classifier, as defined by the DLIB [5] computer vision toolbox.

shape categories, considering that statistically poorly represented classes can be merged with more dominant ones.

The authors of [1] present a novel idea for face shape classification based on three techniques: facial region similarity, correlation and fractal dimensions. Their experiments demonstrate that the proposed approach based on the first technique, namely facial region matching gives effective results for face shape classification. It relies on determining the intersection over union (IOU) between the contour of a human subject's face and an idealistic version of each of the face shape classes. In [8], the authors propose a full pipeline which takes data in the form of a neutral pose image of a female subject, passes it through a classifier to obtain a good estimation of the face shape and finally yields the most appropriate hairstyle recommendation. The core of their pipeline is the VGGNet [12] deep learning classifier architecture, which was successfully combined with feature concatenation and was subjected afterwards to fine-tuning.

The authors of [14] have designed a face shape classifier based on convolutional neural networks (CNN), which they claim is a first in literature. All approaches until their time of writing relied on linear discriminant analysis (LDA), support vector machines (SVM) with different kernel functions, or multi-layer perceptrons (MLP). Their research was driven by the fact that they could refresh these existing techniques using deep learning. Concretely, they employed transfer learning, and retrained the final layer of an Inception v3 architecture [13], thus being able to achieve an accuracy of $\approx 84\%$. Another major contribution from the authors is the creation of their own manually labeled data set, which was made publicly available. Their data consists of 500 images of celebrities for which the face shape is known. There are 100 images per face shape class (heart, oblong, oval, round and square). There is however a trade-off here: there are multiple images of the same celebrity throughout the data set, so instead of having 100 images of distinct individuals per shape class, the data actually contains about 8-10 celebrities per class. Additionally, the authors have tried to bring the subjects in the images to a neutral pose, but one can only cancel the image roll (in the case of 2D images), still leaving the pitch and yaw unresolved.

All the approaches discussed so far are based on 2D images. However, substantially more information regarding the human face shape can be extracted provided we have a full vertex-based model of the face (see Figure 3). Such an approach is discussed in [4], where instead of computing landmark Euclidean 2D distances via a landmark detector, they compute the local deformation of the face in a given basis. They conclude that their proposed method achieves better results than existing methods on extracting the traits of the human face.

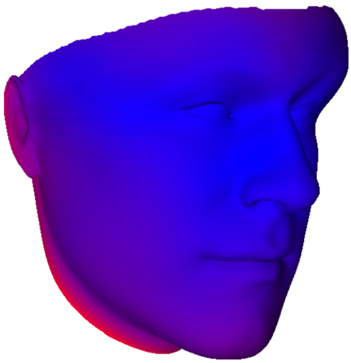

FIGURE 3. Example of 3D heat map visualization of local face vertex deformations versus a standardized, average human face model. A red-shift indicates pronounced deformation, whereas a blue-shift indicates a close match.

3. Proposed approach

The face shape is an important factor in selecting the shape of the eyeglasses; although it is quite difficult to objectively determine the face shape, in the *visagisme* community the following face shapes are generally accepted: rectangle, round, square, heart, diamond, triangle and oval. The rules for determining one's face shape are numerous and leave a lot of space for interpretation, as they involve measuring some features of the face and determining if one measurement is "larger" than another. But all the existing methods rely on measuring the widest part of the face, the height of the face and determining the jaw shape. To automatically measure the face shape we first need to accurately segment the face area and then mimic the measurements required to compute the face shape.

The main difficulty in this task is related to the forehead area, as there are multiple occlusions (hair, bangs, accessories etc.) present in this area and it is quite difficult to determine the boundary between the skin and the hair area. This boundary is required to measure the height of the face (one of the most discriminative measurements when deciding upon the face shape), as well as for the forehead width measurement.

For the shape segmentation we used the same U-Net architecture [9] employed to segment the hair area, as described above, with an off-the-shelf facial landmark detector. To estimate the area of the lower face region, we combined the output of the DLIB [5] facial landmark detector with the segmentation mask. For the forehead estimation, we selected 5 boundary points on the hair segmentation mask, and estimated a symmetrical contour, as depicted in Figure 4.

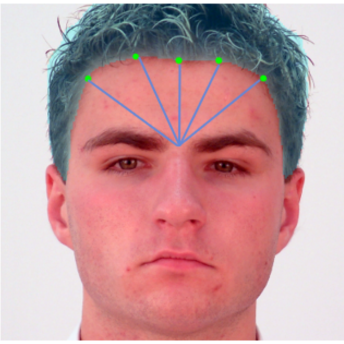

FIGURE 4. The 5 boundary points on the hair segmentation mask, spaced equally at 30 degree angles: 30, 60, 90, 120, 150 degrees, respectively.

The procedure of face segmentation has been already approached and documented, and is available to the general public in the package provided by [5]. On top of this, in order to improve the quality of the classification, we bring our original contribution which derives from a deep learning approach for hair segmentation [2], out of which the forehead line can be extracted, and the facial contour now becomes complete. Finally we extrapolate the full contiguous face contour by merging the face and hair segments, by means of a construct known as *line iterator* (please refer to Figure 5).

Figure 5. The contiguous face contour obtained from the merged face and hair masks.

We determined a set of common, relevant landmarks and measurements that should be used in the classification process. These landmarks are pinpointed in Figure 6: (1) a point in middle of the forehead area, (2) and (10) two extreme points situated to the left and right of the middle forehead point, (3) and (9) two points that determine the largest width of the face, (4) and (8) two points around the jaw, (5) and (7) two points that determine the chin width and (6) the lowest middle point of the chin.

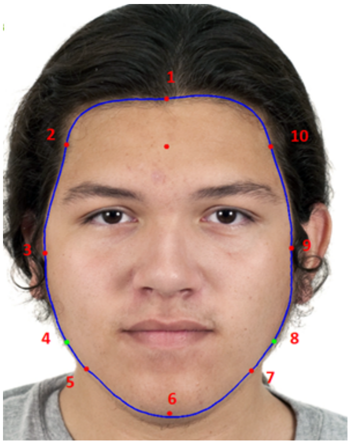

Figure 6. The landmarks used in face shape classification.

The metrics of interest from the contour and internal facial landmarks are listed below:

- face rectangularity; this relies on the minimum bounding rectangle (MBR); the MBR is a standard relationship used to measure the rectangularity of a shape, and it is defined as the ratio of the area

of a region to its minimum bounding rectangle [10]; the face MBR is obtained by computing the MBR for the entire face contour
- middle face rectangularity; the MBR of the contour determined by the points (3), (4), (8) and (9)
- forehead rectangularity; the MBR of the contour determined by the points (3), (9) and (1)
- the chin angle, measured between the points (5), (6) and (7)
- ratio between the lower face width over the middle face width ($RBot$)
- ratio between the upper face width over the middle face width ($RTop$)
- the difference between $RTop$ and $RBot$
- the ratio between the width and the height of the face (fAR)

4. Experimental results

One of the most popular classifiers and also one of the fastest to prototype and train is the naive Bayes classifier [11]. It owes its simplicity to the assumption that every pair of features to be classified is independent of each other. Experimentally, we train a naive Bayes classifier by starting from the Chicago [6] face database (annotated with the face shape tag), on top of which we add 290 images (a morph between existing contour and the corresponding contour template for each face shape type). The details regarding the employed data subset are as follows: **604** total train data set samples, **115** total test data set samples, with a train/test scheme of **85/15**, for which the naive Bayes classifier yields an accuracy of **85%**. As a post-processing step, after we obtain the decision from the naive Bayes classifier, we apply the following post rules of classification, obtained through *empirical experimentation* (here $class_1$ is the class predicted with the highest probability, and $class_2$ is the class predicted with the second-highest probability, respectively). This has been done in an attempt to rectify the misclassification of outliers. Ideally, provided a consistent and balanced data set, these rules should be reconsidered.

```
if class_1 is "Square" and class_2 is "Rectangle" and width / height > 0.75
    then "Rectangle"
if class_1 not "Round" and class_2 is "Square" and width / height > 0.75
    then "Square"
if class_1 is "Oval" and class_2 is "Round" and width / height > 0.75
    then "Round"
if class_1 is "Oval" and class_2 is "Rectangle" and forehead MBR > 0.85
    then "Rectangle"
if class_2 is "Triangle" and RBottom - RTop > 0.10
    then "Triangle"
```

The naive Bayes classifier was one of the candidates for our training, the other being support vector machines (SVMs). In Tables 1 and 2 we give a comparison between the naive Bayes classifier and the SVM classifier, on our data set of choice.

Naive Bayes

Shape	Precision	Recall	F1-score	Support
diamond	1.00	0.60	0.75	5
heart	0.92	0.75	0.83	16
oval	0.86	0.90	0.88	21
rectangle	0.87	0.93	0.90	29
round	0.78	0.88	0.82	16
square	0.83	0.88	0.86	17
triangle	0.80	0.73	0.76	11
accuracy			**0.85**	115

TABLE 1. Results for the naive Bayes classifier, following training and testing. The accuracy *without* the post-processing step is **0.83**.

SVM

Shape	Precision	Recall	F1-score	Support
diamond	0.60	0.60	0.60	5
heart	0.85	0.69	0.76	16
oval	0.67	0.76	0.71	21
rectangle	0.81	0.86	0.83	29
round	0.75	0.75	0.75	16
square	0.93	0.82	0.87	17
triangle	0.55	0.55	0.55	11
accuracy			**0.76**	115

TABLE 2. Results for the support vector machine classifier, following training and testing. The accuracy *without* the post-processing step is **0.73**.

Although the SVM hyper-parameters (*kernel type*, with choices between "*linear*", "*poly*" - polynomial, "*rbf*" - radial basis function or "*sigmoid*"; regularization parameter - "C" and kernel coefficient - "*gamma*") were thoroughly explored using a grid search of available values, still the naive Bayes classifier proves significantly more accurate and was the preferred choice during the face shape application deployment.

As far as future development is concerned, we target the creation of a unified dataset and benchmark for face shape classification, since this is the most important milestone in achieving accurate face shape classification. Currently, in our setup, we hand-picked and manually annotated images which we considered to be representative for their corresponding face shape class. This was done to minimize the bias between two closely related face shapes (such as "diamond" and "heart") and to enforce the robustness of the naive Bayes classifier. In the future we wish to augment our data set with the one supplied by [14].

5. Conclusions

The currently available face shape estimation module makes several assumptions: first of all, it assumes that the person depicted in the image has a near frontal pose. Secondly, as it relies on images, implying 2D projections of the human face, it is quite difficult to extract information about the depth related measurements, such as the length of the jawline. To address this issue, we plan to develop a library to compute a 3D model of the subject's face. We will insist two approaches: one relying on multi-view geometry, while the other using LIDAR data. Once the model of the face is precisely extracted, we can measure all the required distances and angles directly on the 3D model, and therefore develop a classical rule-based algorithm.

Although extracting a 3D face model to estimate the face of the subject can lead to the development of a simple rule based face shape determination algorithm, the problem is that the rules used in face shape determination are highly subjective. Therefore, we envision developing a graph based convolutional neural network model to analyse the relationships between all the relevant facial landmarks and to automatically recognize the face shape.

References

[1] Bansode, N., and Sinha, P. Face shape classification based on region similarity, correlation and fractal dimensions. *IJCSI International Journal of Computer Science Issues 13*, 1 (2016).

[2] Ileni, T. A., Borza, D. L., and Darabant, A. S. Fast in-the-wild hair segmentation and color classification. In *Visigrapp* (2019).

[3] Ion Marinescu, A., Alexandru Ileni, T., and Sergiu Dărăbant, A. A versatile 3d face reconstruction from multiple images for face shape classification. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (Sep. 2019), pp. 1–6.

[4] Jiang, Z., Wu, Q., Chen, K., and Zhang, J. Disentangled representation learning for 3d face shape. *CoRR abs/1902.09887* (2019).

[5] Kazemi, V., and Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2014).

[6] Ma, D., Correll, J., and Wittenbrink, B. The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods 47* (01 2015).

[7] Marinescu, A. I., Dărăbant, A. S., and Ileni, T. A. A fast and robust, forehead-augmented 3d face reconstruction from multiple images using geometrical methods. In

2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (Sep. 2020), pp. 1–6.

[8] Pasupa, K., Sunhem, W., and Chu Kiong, L. A hybrid approach to building face shape classifier for hairstyle recommender system. *Expert Systems with Applications 120* (11 2018).

[9] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597* (2015).

[10] Rosin, P. Measuring shape: Ellipticity, rectangularity, and triangularity. *Machine Vision and Applications 14* (08 2001).

[11] Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*, 3 ed. Prentice Hall, 2010.

[12] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015), Y. Bengio and Y. LeCun, Eds.

[13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. *CoRR abs/1512.00567* (2015).

[14] Tio, A. E. D. Face shape classification using inception v3. In *Electrical and Electronics Engineering Institute, University of the Philippines Diliman, Quezon City, Philippines* (2017).

Dept. of Computer Science, Babeș-Bolyai University, Kogălniceanu 1, 400084 Cluj-Napoca, Romania

Email address: `alexandru.marinescu@ubbcluj.ro`