Comparison and Fusion of Multiple Types of Features for Image-Based Facial Beauty Prediction

Fangmei Chen[1(✉)], David Zhang[2], Cunrui Wang[3], and Xiaodong Duan[3]

[1] Information and Communication Engineering Department, Dalian Minzu University, Dalian 116600, China
cfm@dlnu.edu.cn

[2] Biometrics Research Centre and the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[3] Computer Science and Engineering Department, Dalian Minzu University, Dalian 116600, China

Abstract. Facial beauty prediction is an emerging research topic that has many potential applications. Existing works adopt features either suggested by putative rules or borrowed from other face analysis tasks, without an optimization procedure. In this paper, we make a comprehensive comparison of different types of features in terms of facial beauty prediction accuracy, including the rule-based features, global features, and local descriptors. Each type of feature is optimized by dimensionality reduction and feature selection. Then, we investigate the optimal fusion strategy of multiple types of features. The results show that the fusion of AAM, LBP, and PCANet features obtains the best performance, which can serve as a competitive baseline for further studies.

Keywords: Facial beauty · Feature extraction · Fusion

1 Introduction

Image-based facial beauty prediction has many potential applications such as aesthetic surgery planning, cosmetic recommendation, photo retouching, entertainment, etc. The core of facial beauty prediction is discovering the relationship between low-level visual features and high-level perceived attractiveness.

Two categories of features have been adopted in existing works. One includes putative rules, most of which are defined in the form of some ideal ratio. Computational models of facial beauty have been built with those putative ratio features [1–3]. It is found that only a small subset of ratio features is important for facial beauty prediction [2]. Another category of features is inspired from face recognition studies, e.g., shape parameters, eigenface, Gabor filter responses [4], local binary patterns (LBP) [5], etc. Researchers often combine multiple types of facial features to build regression models of facial beauty. For example, Eisenthal et al. [6] combine geometric features, hair color, and skin smoothness into the

J. Zhou et al. (Eds.): CCBR 2017, LNCS 10568, pp. 23–30, 2017.
https://doi.org/10.1007/978-3-319-69923-3_3

regression model. Nguyen et al. [7] concatenate LBP, Gabor filter responses, color moment, shape context, and shape parameters as a feature vector and apply PCA to reduce the dimensionality. Gray et al. [8] design a multi-scale feature model by local filters and down-sampling. Although so many features have been used for facial beauty prediction, few works compare the discriminative power of different types of features and investigate the optimal fusion strategy.

In this paper, we make a comprehensive study on comparison and fusion of features for facial beauty prediction. Besides the features used in previous works, AAM parameters [9] and PCANet features [10] are also employed. Firstly, different types of features are compared in terms of facial beauty prediction performance. Secondly, dimensionality reduction and feature selection techniques are performed to optimize each type of feature. Thirdly, we investigate the optimal fusion strategy of multiple types of features to further improve the performance. The model built by the optimal fusion strategy can serve as a competitive benchmark for further studies.

2 Feature Extraction

Most putative rules on facial beauty are defined as ratios. We review the relevant studies and provide a concise summary in our previous work [18]. There are 26 putative ratio rules, including the neoclassical and golden ratio rules [1,2]. 98 landmarks are extracted (see Fig. 1), which are distributed on the major facial organs. Given the landmarks, it is convenient to calculate the ratios.

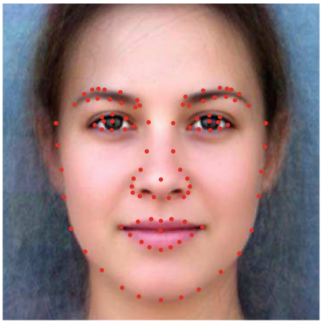

Fig. 1. The layout of 98 landmarks on a face image.

For global features, shape parameters, eigenface, and AAM features are considered. First, the shape, texture, and appearance models are learned on a training face image set. Then, each input face image can be represented by a vector of model parameters. The x- and y- coordinates of the 98 landmarks are concatenated to form a vector that represents the geometry of the face, i.e., $[x_1, x_2, \ldots, x_{98}, y_1, y_2, \ldots, y_{98}]^T$. Given the landmark vectors, shape features can be obtained by Procrustes superimposition [11]. Then, PCA is performed on the shape vectors to reduce the dimensionality to 50, which keeps 97.3% of the total energy. The 50-dimensional component scores are called shape parameters. Eigenface [12] is a method of texture representation. A face image is represented by a vector of pixel values. Suppose the training data are $I_1, \ldots, I_n$, the mean vector μ is subtracted from I_i, and a data matrix $X = [I_1 - \mu, \ldots, I_n - \mu]$ is constructed. The eigenvectors of $X^T X$ are called eigenfaces. AAM [9] parameterizes and combines the shape and the texture of a face. We keep the first 100 dimensions, which explains 97% of the appearance variations.

Local descriptors are also considered. Gabor filters encode facial shape and texture information over a range of spatial scales. Following [4], we use five scales and eight orientations. Other parameters are set as $k_{max} = \pi$, $f = \sqrt{2}$, and $\sigma = 2\pi$. LBP operator encodes every pixel of an image with an 8-bit binary number by thresholding the neighborhood of the pixel with the center pixel value. The histogram of the labels is used as a texture descriptor. In our implementation, the face images are cropped and resized into 128×128 and divided into 7×7 local regions. Then the $\mathrm{LBP}_{8,2}^{u2}$ operator [5] is applied to each region. PCANet is a simple deep learning network proposed by [10]. It has three components: cascaded PCA filtering, binary hashing, and block-wise histogram, as shown in Fig. 2. For each input image, the output of the first layer filtering is L_1 images, which are inputs for the second layer. Each of the L_1 input images has L_2 outputs, which are binarized and converted to a single integer-valued image by

$$I^{(3)} = \sum_{l=1}^{L_2} 2^{l-1} I_l^{(2)}. \tag{1}$$

The pixel value of $I^{(3)}$ is in the range $[0, 2^{L_2} - 1]$. The L_1 images in layer three are divided into 64 blocks of size 16×16. The histograms of the blocks are cascaded to obtain the final feature vector. In our implementation, $k_1 = k_2 = 7$, $L_1 = L_2 = 8$.

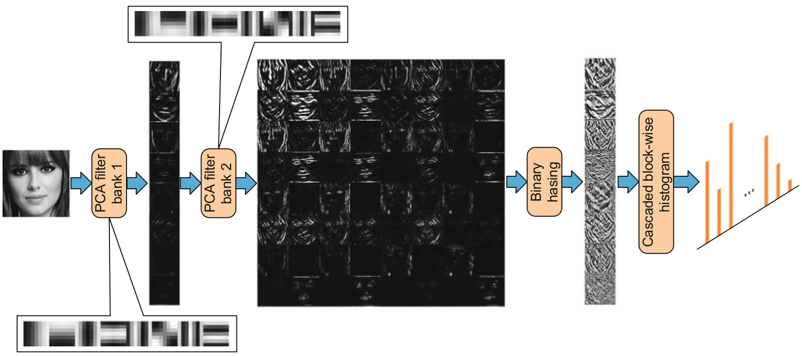

Fig. 2. Illustration of PCANet feature extraction.

3 Feature Selection

Because the number of training data is often limited, removing the irrelevant and redundant variables will alleviate the curse of dimensionality. We adopt the lasso [13] method for feature selection for its good performance and moderate computational cost. The objective function of the lasso is

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}}\{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\beta\|_2^2 + \lambda\|\beta\|_1\}, \tag{2}$$

where $\boldsymbol{X}$ is the input data matrix, $\boldsymbol{y}$ is the beauty score vector, and λ is the regularization parameter. The l_1 penalty will promote sparse solutions. To solve the entire lasso path, we use SpaSM, a toolbox for sparse statistical modeling [14]. The variables are then sorted according to the path.

Given the sorted variables, a nested sequence of models can be obtained. To decide the optimal number of variables to keep, we train models with increasing number of sorted variables. The first k variables are selected if none of the larger feature subsets can increase the prediction accuracy significantly, which is determined by two-sample t-tests.

4 Optimal Fusion Strategy

In this section, we investigate the combinations of multiple types of features to further increase the facial beauty prediction accuracy. Seven types of features are introduced in Sect. 2. Hence, there are totally 127 combinations, which can be represented by $Comb = [C_1, C_2, \dots, C_{127}]$, where $C_1 = \{1\}$, $C_2 = \{1, 2\}, \dots, C_{127} = \{1, 2, 3, 4, 5, 6, 7\}$. They are evaluated by score level fusion performance. The outputs of the seven models are denoted by $y_1, \dots, y_7$. For the combination C_k, the score level fusion result is

$$\hat{y} = \sum_{i \in C_k} w_i^{(k)} y_i. \tag{3}$$

The weight vector $\boldsymbol{w}^{(k)}$ can be obtained by solving the least squares problem

$$\boldsymbol{w}^{(k)} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \|\boldsymbol{y} - Y^{(k)}\boldsymbol{w}\|, \tag{4}$$

where $Y^{(k)}$ is a $n \times |C_k|$ matrix including n entries of model outputs selected by C_k, and $\boldsymbol{y}$ is human-rated beauty scores. The correlation between the predicted scores $\hat{\boldsymbol{y}}$ and human rated scores $\boldsymbol{y}$ is used to evaluate the prediction performance, i.e.,

$$r_{\boldsymbol{y},\hat{\boldsymbol{y}}} = \frac{cov(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\sigma_{\boldsymbol{y}}\sigma_{\hat{\boldsymbol{y}}}} = \frac{E[(\boldsymbol{y} - \mu_{\boldsymbol{y}})(\hat{\boldsymbol{y}} - \mu_{\hat{\boldsymbol{y}}})]}{\sigma_{\boldsymbol{y}}\sigma_{\hat{\boldsymbol{y}}}}. \tag{5}$$

The optimal fusion strategy is the least complex model with the most competitive performance.

5 Experiments

5.1 Data Set and Preprocessing

The experiments are based on the database built by [15], which includes 390 celebrity face images of Miss Universe, Miss World, movie stars, and super models collected from the web and 409 common face images. The beauty scores are given. Active shape model (ASM) [16] is used to detect the landmarks, and we develop a tool for manual adjustment to further improve the precision of landmark positions. To extract eigenface, Gabor, LBP, and PCANet features, the images are cropped with a squared bounding box and resize the cropped image into 128×128.

5.2 Comparison of Features

The features are compared in terms of facial beauty prediction performance, which is measured by the correlation between predicted beauty scores and human-rated scores. Three statistical regression methods were applied to build the computational models: K-nearest neighbor (KNN), linear regression (LR), and support vector regression (SVR) [17]. In KNN regression, the parameter k was optimized by search. The SVR method used RBF kernels, and the parameters were obtained by grid search. As the original Gabor, LBP, and PCANet features are high-dimensional, PCA was performed to reduce the dimension to 100, which is the same with the eigenface and AAM features. We randomly selected 90% of the data for training and the remaining 10% were used for testing. The procedures were run 100 times. Figure 3 shows the average prediction accuracies of the three methods with different types of features. We can see that PCANet achieves the highest accuracy, followed by AAM. Among the three regression methods, SVR achieves the best performance consistently.

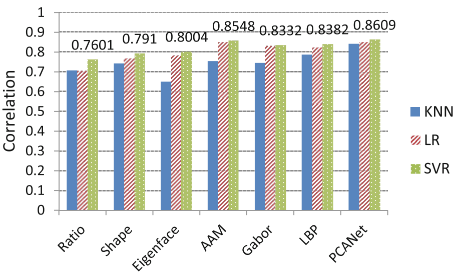

Fig. 3. Comparison of regression methods in terms of facial beauty prediction accuracy.

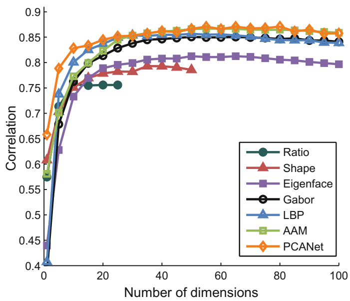

Fig. 4. Prediction performance with increasing number of selected feature dimensions.

5.3 Feature Selection Results

A series of models were trained with increasing number of features ordered by the lasso method. The models were built with SVR method, as suggested by Sect. 5.2. The results are plotted in Fig. 4. We can see that at first the performances of all types of features increase rapidly, and after selecting about 20% of the total features, the performance curves become flat. Hence, there are irrelative or redundant variables. The numbers of selected features are determined by multiple two-sample t-tests with significance level $\alpha = 0.05$. For most of the cases, about a half of features are discarded. However, the performances of the selected features are even slightly better than those of the original features.

5.4 Optimal Fusion Strategy

In this part, we investigated the optimal fusion strategy of multiple types of features. For 7 types of features, there are 127 different combinations. In order to obtain the combination weights defined in (4), we need to construct the single-type-feature output matrix Y. We randomly selected 90% of the faces to train 7 models corresponding to the 7 types of features. The models predicted beauty scores of the remaining 99 faces, so that there were 99×7 scores. Repeating this procedure for 10 times, we obtained a matrix of size 990×7, which served as the matrix Y. Then, by solving (4), we got the combination weights. The final scores are weighted sum of the outputs of the n models. We run 10-fold cross validation and the average performances of different fusion strategies are plotted in Fig. 5. The optimal strategy is the most parsimonious one with competitive prediction accuracy, as marked in Fig. 5. Table 1 shows the optimal combinations constrained by the number of feature types. We can see that the fusion of PCANet, LBP, and AAM can significantly improve the prediction performance, and adding more types of features cannot increase the performance significantly.

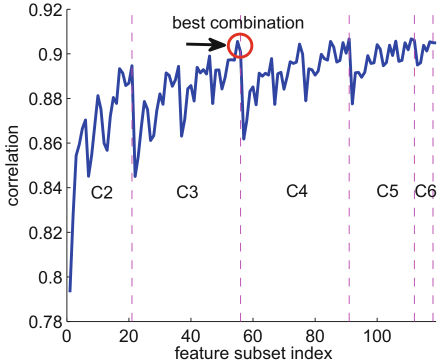

Fig. 5. Prediction performance of different feature combinations. For example, C2 means all combinations of two feature types, which has $C_7^2 = 21$ subsets.

Table 1. Optimal combinations and score level fusion results

No. of types	Optimal combination	Performance
1	PCANet	0.8708
2	PCANet+LBP	0.8947
3	PCANet+LBP+AAM	0.9056
4	PCANet+LBP+AAM+Gabor	0.9067

5.5 Comparison with Other Works

Our method was compared with those proposed by other works. As facial beauty prediction is still an emerging topic, there is no public database for comparison. We implemented the methods of other works on our database. As shown in Table 2, three methods are compared. [2] build a linear model with ratio features. [8] train a convolutional neural network for facial beauty prediction. [7] cascade shape, Gabor, LBP, and color moment features into a high-dimensional vector and performs PCA to reduce the feature dimension to 350. We did not include the color moment feature, because the images were collected from many sources, and the various illumination conditions distort the true face colors. For each method, 10-fold cross-validation was performed and the average prediction accuracy is shown in Table 2. The results show that our method is much better than other methods.

Table 2. Comparison with other works

Work	Feature	#Dimension	Method	Performance
[2]	Ratios	23	LR	0.6958
[8]	convNet	-	Neural network	0.7512
[7]	Shape+Gabor+LBP	350	SVR	0.8185
Our	AAM+LBP+PCANet	151	SVR+fusion	**0.9056**

6 Conclusion

In this paper, we give a comprehensive study on feature design and fusion strategy for facial beauty prediction. Ratio, shape, eigenface, AAM, Gabor, LBP, and PCANet features are compared in terms of facial beauty prediction accuracy. It is found that feature selection can promote the prediction performance. For feature selection, the lasso method performs better than filter and sequential wrapper methods. Then the optimal fusion strategy of multiple types of features is investigated. The results show that the best single-type feature is PCANet, and the optimal feature combination is PCANet, LBP, and AAM. By score level fusion, the final model achieves a correlation of 0.9056 between predicted scores and human rated scores, which is much better than existing methods.

References

1. Gunes, H., Piccardi, M.: Assessing facial beauty through proportion analysis by image processing and supervised learning. Int. J. Hum. Comput. Stud. **64**(12), 1184–1199 (2006)
2. Schmid, K., Marx, D., Samal, A.: Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios. Pattern Recogn. **41**(8), 2710–2717 (2008)
3. Fan, J., Chau, K.P., Wan, X., Zhai, L., Lau, E.: Prediction of facial attractiveness from facial proportions. Pattern Recogn. **45**(6), 2326–2334 (2012)
4. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Process. **11**(4), 467–476 (2002)
5. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
6. Eisenthal, Y., Dror, G., Ruppin, E.: Facial attractiveness: beauty and the machine. Neural Comput. **18**(1), 119–142 (2006)
7. Nguyen, T.V., Liu, S., Ni, B., Tan, J., Rui, Y., Yan, S.: Towards decrypting attractiveness via multi-modality cues. ACM Trans. Multimedia Comput. Commun. Appl. **9**(4), 28 (2013)
8. Gray, D., Yu, K., Xu, W., Gong, Y.: Predicting facial beauty without landmarks. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 434–447. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15567-3_32
9. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
10. Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: PCANet: A Simple Deep Learning Baseline for Image Classification? arXiv:1404.3606 (2014)
11. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. Wiley, Hoboken (1998)
12. Sirovich, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. JOSA A **4**(3), 519–524 (1987)
13. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. SSS. Springer, New York (2009). doi:10.1007/978-0-387-84858-7
14. Sjöstrand, K., Ersbøll, B.: SpaSM: A MATLAB Toolbox for Sparse Statistical Modeling, October 2015. http://www2.imm.dtu.dk/projects/spasm/
15. Chen, F., Xu, Y., Zhang, D.: A new hypothesis on facial beauty perception. ACM Trans. Appl. Percept. **11**(2), 1–20 (2014)
16. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
17. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995). doi:10.1007/978-1-4757-2440-0
18. Chen, F., Zhang, D.: Evaluation of the putative ratio rules for facial beauty indexing. In: 2014 International Conference on Medical Biometrics, pp. 181–188. IEEE Press, New York (2014)