

Article

A Novel Framework for Assessing Facial Attractiveness Based on Facial Proportions

Yu-Jin Hong ^{1,2}, Gi Pyo Nam ², Heeseung Choi ² , Junghyun Cho ²  and Ig-Jae Kim ^{1,2,*}

¹ Department of HCI Robotics, University of Science and Technology, Daejeon 34113, Korea; g10518@kist.re.kr

² Center for Imaging Media Research, Korea Institute of Science and Technology, Seoul 02792, Korea; gpnam@kist.re.kr (G.P.N.); hschoi@kist.re.kr (H.C.); jhcho@kist.re.kr (J.C.)

* Correspondence: drjay@kist.re.kr; Tel.: +82-2-958-5766

Received: 11 October 2017; Accepted: 26 November 2017 ; Published: 29 November 2017

Abstract: In this paper, we present a novel framework for automatically assessing facial attractiveness that considers four ratio feature sets as objective elements of facial attractiveness. In our framework, these feature sets are combined with three regression-based predictors to estimate a facial beauty score. To enhance the system's performance to make it comparable with human scoring, we apply a score fusion technique. Experimental results show that the attractiveness score obtained by the proposed framework better correlates with human assessments than the scores from other predictors. The framework's modularity allows any features or predictors to be integrated into the facial attractiveness measure. Our proposed framework can be applied to many beauty-related fields, such as the plastic surgery, cosmetics, and entertainment industries.

Keywords: facial attractiveness; facial beauty; ratio feature; score fusion

1. Introduction

Attractiveness is one of the most critical social characteristics of the human face. Several studies have shown that facial attractiveness is considered to be a basis for social and intellectual competencies, such as mate choice and income [1,2]. In recent years, beauty-related industries have grown rapidly in many countries [3]. Thus, the analysis and measurement of facial attractiveness have garnered attention from scientists, physicians, and artists because of its many potential applications in entertainment, virtual media, plastic surgery, and cosmetic industries. Although it is debatable whether facial attractiveness is objective or subjective, recent empirical results support the idea that attractiveness is objective, achievable by measuring cross-cultural consistency [4], brain activity patterns [5], and infant preference [6]. Studies in psychology and medical science have also explored aesthetic evidence of what makes a face more attractive, finding facial averageness [7] and symmetry [8] to be important. That is, facial attractiveness increases as averageness and symmetry increase. Facial skin colors and texture are also significantly related to face attractiveness [9,10], and has been included in evaluations of facial attractiveness [11,12]. Moreover, research has shown that attractive faces should follow certain defined ratios of facial proportions, such as neoclassical canons [13] and the golden ratio [14] which have been believed to be ideal ratios for beautiful faces since ancient times by artists, physicians, and orthodontists [15]. In machine learning, several methods have been proposed to assess facial attractiveness by encoding these objective factors from a face. However, due to the improper feature extraction, these elements have not been investigated thoroughly. Moreover, they have not been compared and combined in an overall framework, even though this approach might decrease the gap between human and machine performance. Therefore, there is a need to develop an efficient framework as an initial step towards a machine obtaining human-level performance in predicting facial attractiveness. In this paper, we focus on developing a framework for assessing

facial attractiveness based on facial proportion factors widely believed to be objective elements for facial beauty. The remainder of the paper is organized as follows. Section 2 presents a brief summary of previous machine learning approaches for assessing or enhancing facial attractiveness. Section 3 explains the proposed framework. Section 4 discusses the experimental results and possible applications. Conclusions and future work are addressed in Section 5.

2. Related Work

Researches in psychology and biology settle this problem by hypothesizing which aspects of faces make them attractive. Studies indicate that features such as symmetry, averageness, and sexual dimorphism influence the perception of attractiveness. Jones [12] demonstrated that healthy-looking women appear more attractive based on these three features. Galton [16] proved that average faces have significant correlation with facial beauty, whereas very attractive faces are not average [17]. Based on facial averaging, face symmetry indeed increases, which supports evolutionary biology studies that indicate the positive influence of facial symmetry on attractiveness [18]. Further, sexual dimorphism—the secondary sexual characteristics appearing during puberty can affect attractiveness judgments [19]. Specifically, different sexually dimorphic features for males and females make them appear more masculine or feminine. Many studies have provided evidence that masculinity and femininity are more convincing attractiveness features than symmetry [12,20,21]. Skin color and texture are also known to be related to attractiveness. As many researchers have proposed a link between attractiveness and traits that appear healthy, the health of facial skin might be a surface property that positively influences attractiveness judgments. Fink et al. [9] evaluated facial beauty via human scoring from skin colors and textures. In Jones et al. [22], facial skin is correlated with male facial attractiveness and is a visual cue for attractiveness judgments. Since facial attractiveness is influenced by a range of factors, studies have considered both facial shape and appearance for facial attractiveness judgments. Kagan et al. [11] analyzed facial attractiveness based on facial appearance and geometry. Bronstad and Russell [21] described the effect of facial shape and appearance (texture) on attractiveness measurements using pixel information in an image, and confirmed that the two criteria had similar effects. Jones [12] also evaluated facial attractiveness by landmark-based feature distance (facial shape) and skin color and showed that both facial shape and skin characteristics significantly affect facial beauty measurements.

Recently, machine learning approaches have been developed to assess facial attractiveness. Aarabi et al. [23] developed a preliminary automatic facial beauty scoring system based on a vector of eight ratio values between facial landmarks (eyes, eyebrows, mouth and nose). The 12 human assessors scored 40 training images on a four-point scale. To estimate a test image's score, its feature vector was computed and then the average score of the 10 nearest faces in the training set was calculated. Gunes and Piccardi [24] collected 215 female face images and the images which were rated on a 10-point scale by 48 human assessors. A vector of 13 ratio values, including the golden ratio, was calculated between facial landmarks. Then a trained, tree-based classifier was used to determine the beauty score of the test images. Kagan et al. [11] extended their earlier work [25] to develop a regression-based facial attractiveness predictor. They adopted both geometric distance features and skin color features to estimate facial attractiveness. From the 84 facial landmarks, 3486 distance values were calculated; these were reduced to 90 values by principal component analysis (PCA). The eight feature components calculated from skin and hair color values were combined with these 90 features. A total of 98 features were used by three predictors (linear regression, a support vector machine, and a Gaussian process). The results of rating 91 frontal face images on a seven-point scale showed that the system with the best predictor (linear regression) had a Pearson correlation of 0.82 with ratings by 27 human raters. A similar regression-based approach [26] was proposed to determine the relation between three facial proportion factors—the neoclassical canon, symmetry and the golden ratio—with human ratings. From the face recognition technology (FERET) database, 420 frontal face images were selected and scored on a 10-point scale by 36 human assessors. These three feature sets

were extracted from 29 facial landmarks and statistical analysis software was used to calculate the correlation between the machine and human ratings. Unfortunately, these predictors had little relation to beauty scores. However, motivated by the work [11], a data-driven approach to enhancing facial attractiveness was proposed in [27]. Similarly, 234 distance feature components were calculated from 84 facial landmarks. Then, with a trained beautification engine based on support vector regression (SVR) or a K-nearest neighbor (K-NN) algorithm, the system searched the face space for a nearby point with a higher predicted attractiveness rating. Finally, the triangulation of the original face is warped toward those of the beautiful faces more similar to the original one. In recent years, attractive measurement and evaluation of 3D faces have been introduced because they produce more precise and accurate prediction results. O'Toole et al. [28] calculated average face shapes and textures based on 3D scan data of 200 men and women, and proved that averageness influences attractiveness in 3D faces, just as in 2D faces. Fink et al. [29] revealed that skin texture has a considerable influence on the attractiveness evaluation of 3D faces by constructing a skin map with various skin colors that is fitted to a 3D face template to normalize face shape. Jang et al. [30] estimated facial attractiveness through manually detected landmarks on a 3D face, and calculated the height of the nose, the horizontal and vertical curvatures of the forehead, the curvature of the cheek, and the chin volume. They showed that more attractive faces have features such as more protruding noses with greater nasolabial angles and greater vertical curvature of the forehead. They also described that more attractive faces are highly correlated with the beauty scores measured by the neoclassical canon and symmetry. Since 3D faces have volume and surface curvature, unlike 2D faces, there is a need to modify steps such as facial landmark extraction and distance calculation to measure the beauty of 3D faces. Vezzetti et al. [31] introduced an automatic landmark extraction method on a 3D face. Marcolin et al. [32] also devised facial geometrical descriptors representing symmetry features for the attractiveness analysis of 3D faces. Liao et al. [33] manually selected landmarks for the golden ratio, neoclassical canon, and symmetry on a 3D face, and these distances and ratio values were revised by considering the surface characteristics of a 3D face. Table 1 summarizes the related studies. In this paper, we assume that facial attractiveness is related to harmonious facial proportions. To prove this assumption, our method is based on the ratio features that reflect several important facial proportions. Moreover, a facial beauty score can be calculated automatically by learning of the relation of ratio features and human ratings.

Table 1. Related work for assessing facial attractiveness.

Dimension	Reference	Feature(s)	Measurement(s)
2D	Aarabi et al. [23]	Facial ratio	8-element vector of ratios
	Gunes and Piccardi [24]	Facial ratio	Golden ratio, facial thirds
	Schmid et al. [26]	Facial ratio	Golden ratio, symmetry, canons
	Perrett et al. [4]	Facial landmarks	Averageness
	Bronstad and Russell [21]	Landmark distance, pixels	Symmetry, averageness, sexual dimorphism
	Kagian et al. [11]	Landmark distance, skin	Geometric features
	Jones [12]	Landmark distance, skin	Symmetry, averageness, femininity, skin tones
	Galton et al. [16]	Shape, pixels	Averageness
	Eisenthal et al. [25]	Shape, hair color	Symmetry
	Jones et al. [22]	Skin	Skin patches
3D	Fink et al. [9]	Skin	Skin colors, texture
	Fink et al. [29]	Skin	Skin colors
	O'Toole et al. [28]	Shape, texture	Shape, texture
	Liao et al. [33]	Facial ratio	Golden ratio, symmetry, canons

3. Proposed Framework

A flowchart of the proposed framework is illustrated in Figure 1. The framework consists training and testing procedures. In training, four facial ratio feature sets (RFSs) were extracted from each training image, along with the average attractiveness score from human raters. The optimal parameter sets of each predictor (based on regression) were obtained for each feature set. In testing, given an arbitrary face image, the trained predictor for each feature set was applied to corresponding

feature sets to obtain scores for each. Then, the final score was fused at the score level. Due to the modular structure of the framework, any features and predictors can be applied. In the remainder of this section, we explain the proposed method in detail.

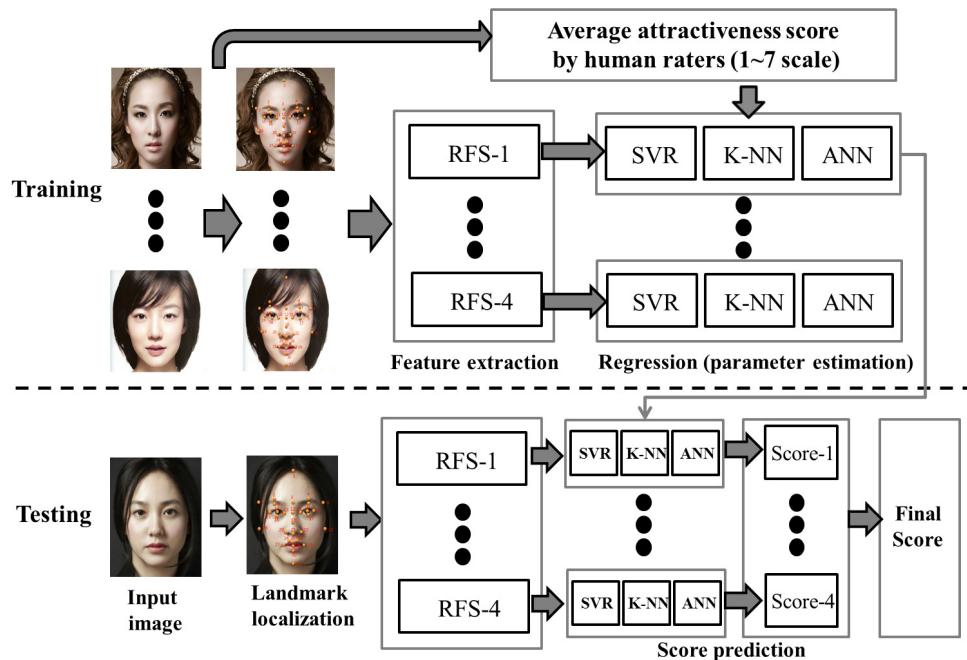


Figure 1. A proposed framework for assessing facial attractiveness.

3.1. Training Procedure

3.1.1. Facial Landmark Localization

To detect facial landmarks automatically, the active shape model (ASM)-based method [34] is applied to each face. The open source software STASM [35] was utilized as an initial detector for 81 predefined facial landmarks. Then, only 31 landmarks (covering the eyebrows, eyes, noses, mouth, forehead, and head contours), sufficient for calculating ratio feature values, were selected (see Figure 2).

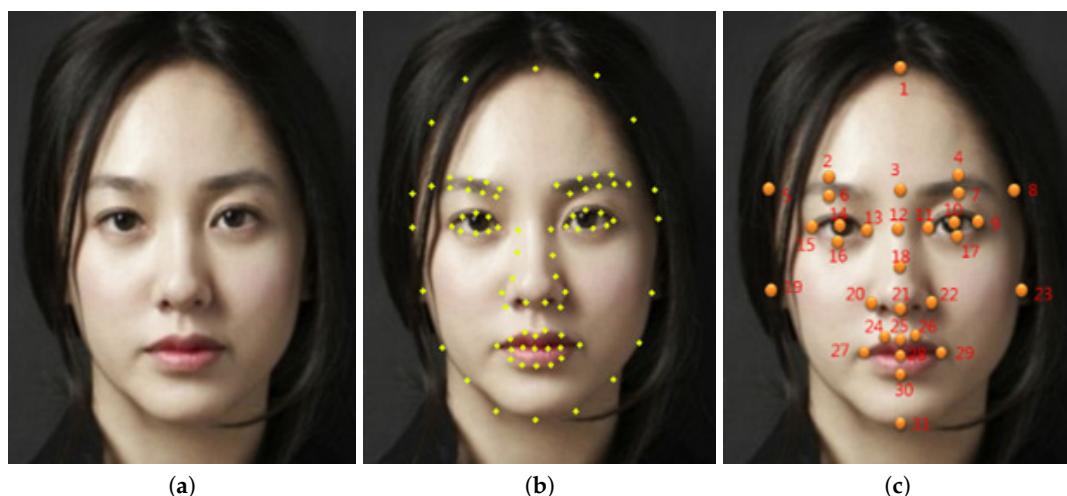


Figure 2. Facial landmark localization: (a) Original face image; (b) 81 facial landmarks found by STASM [35]; (c) Refined set of 31 facial landmarks.

3.1.2. Ratio Feature Set (RFS) Extraction

After acquiring facial landmarks, ratio feature components which have been reported as an objective elements for facial attractiveness are extracted. Neoclassical canon refers to the ratio used by Renaissance artists to draw a beautiful face. Farkas et al. [13] summarized these rules into nine neoclassical features. Neoclassical canonical features were proposed long ago and, despite the changes in beauty standards over time, are still valid measures in the present age in anatomy, art, and medicine studies [36,37]. Although there is controversy over the correlation between the golden ratio and facial beauty [38–41], we tested it as one of our ratio feature sets, because it accounts for the ratio between elements of the eyes, nose, and mouth, as well as the aspect ratio of the width and height of the face. Symmetry is known to be important in evaluating facial attractiveness [12,18], and has been defined and studied in various ways. In many previous studies, symmetry was measured along the vertical axis of the face. In this study, we measured symmetry based on the center point of the face, thus considering both the left and right sides of the face [12,26]. In addition, eight ratio values have been proposed by artists and scientists [23,42–44]. This measure assumes that the facial contour is an ellipse and detects the exact position of the eyes and mouth to calculate eight proportional features, enabling simple measurements with several facial landmarks. Schmid et al. [26] analyzed the role of symmetry, neoclassical canons, and golden ratio for the determination of facial attractiveness. The basic premise is that portions of an attractive face should follow certain defined ratios. They summarized these principles in nine neoclassical canons (RFS-1), 14 golden ratios (RFS-2) and nine symmetry values (RFS-3). Moreover, according to [23], they applied eight ratio values (RFS-4) to determine the facial attractiveness. We call these four types of feature vectors as ratio feature sets (RFSs) hereafter. To assess facial attractiveness based on facial proportion features, we applied nine neoclassical canon (RFS-1), 14 golden ratio (RFS-2), nine symmetry (RFS-3), and eight ratio (RFS-4) values. A detailed description and used landmarks for calculating the feature sets are shown in Table 2 and Figure 3. In Table 2, $\text{dist}(a,b)$ means Euclidean distance of the landmark a and b and $\text{mid}(a,b)$ calculates the average point of the landmark a and b. These RFSs were incorporated as input vectors in our framework. Due to the different value ranges of each feature set, these RFSs were normalized. In the golden ratio set (RFS-2), attractive faces should have ratios approaching a value of 1.618. In the neoclassical canon (RFS-1) and symmetry (RFS-3) sets, the ratios should approach 1. Therefore, to normalize features into a unit interval [0, 1], we applied the exponential function to each ratio value as follows:

$$x_n = e^{-(x-m)^2}, \quad (1)$$

where x and x_n denote the original and normalized ratio values, respectively, and m is 1.618 for RFS-2 and 1 for RFS-1 and RFS-3. Meanwhile, RFS-4 does not follow any predefined ratio, so min-max normalization was performed.

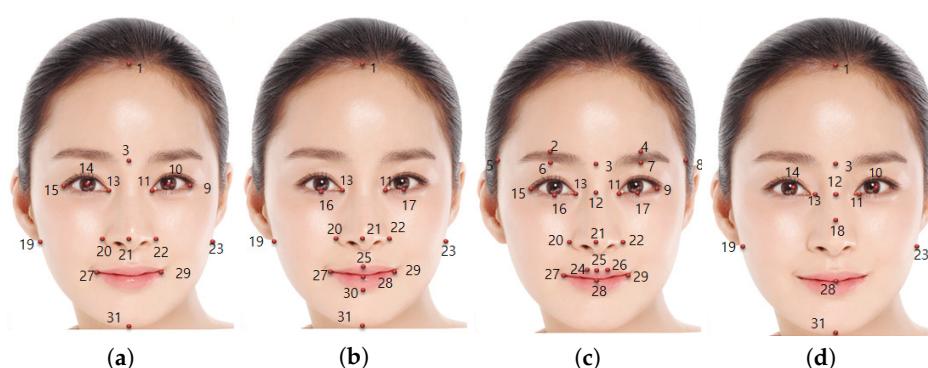


Figure 3. Landmarks for ratio feature sets (RFSs). (a) RFS-1: Neoclassical canons; (b) RFS-2: Golden ratio; (c) RFS-3: Symmetry; (d) RFS-4: Eight ratio values.

Table 2. Descriptions of the four feature sets (RFSs).

RFS-1: Neoclassical Canons		RFS-2: Golden Ratio	
Description (Distance/Distance)	Ratio vector	Description (Distance/Distance)	Ratio vector
facetop-eyebrows/eyebrows-nose	dist(1,3)/dist(3,21)	under eyes/interocular	dist(16,17)/dist(13,11)
eyebrows-nose/nose-jaw	dist(3,21)/dist(21,31)	under eyes/nose width	dist(16,17)/dist(20,22)
facetop-eyebrows/nose-jaw	dist(1,3)/dist(21,31)	mouth width/interocular	dist(27,29)/dist(13,11)
interocular/nose width	dist(11,13)/dist(20,22)	upper lip-aw/interocular	dist(25,31)/dist(13,11)
interocular/pupil-outer eye(right)	dist(11,13)/dist(13,15)	upper lip-jaw/nose width	dist(25,31)/dist(20,22)
interocular/pupil-outer eye(left)	dist(11,13)/dist(9,11)	interocular/lip height	dist(13,11)/dist(25,30)
right eye width/left eye width	dist(13,15)/dist(9,11)	nose width/interocular	dist(20,22)/dist(13,11)
mouth width/(1.5 × nose width)	dist(27,29)/(1.5 × dist(20,22))	nose width/upper lip height	dist(20,22)/dist(25,28)
face width/(4 × nose width)	dist(19,23)/(4 × dist(20,22))	interocular/nose-mouth	dist(13,11)/dist(21,28)
RFS-3: Symmetry		lip height/nose-mouth	dist(25,30)/dist(21,28)
Descriptions(Distance/Distance)	Ratio vector	face height/face width	dist(1,31)/dist(19,23)
Eyebrows	dist(2,3)/dist(4,3)	nose-jaw/mouth-jaw	dist(21,31)/dist(28,31)
	dist(6,3)/dist(7,3)	nose width/nose-mouth	dist(20,22)/dist(21,28)
Eyes	dist(6,3)/dist(7,3)	mouth width/nose width	dist(27,29)/dist(20,22)
	dist(11,12)/dist(13,12)	RFS-4: 8-Ratio Vectors	
Nose	dist(16,12)/dist(17,12)	Descriptions (Distance/Distance)	Ratio vector
	dist(20,21)/dist(22,21)	interpupillary/eyes-mouth	dist(10,14)/dist(12,28)
Mouth	dist(24,21)/dist(26,25)	interpupillary/face width	dist(10,14)/dist(19,23)
	dist(27,28)/dist(27,29)	eyes-mouth/face height	dist(12,28)/dist(1,31))
Face	right pupil-mouth/ left pupil-mouth	right pupil-mouth/ left pupil-mouth	dist(14,28)/dist(10,28)
	dist(5,3)/dist(8,3)	facetop-eyebrows/eyes-jaw	dist(1,3)/dist(12,31)
	nose-mouth/mouth-jaw	nose-mouth/mouth-jaw	dist(12,28)/dist(1,31)
	eyes center-nose/nose-jaw	eyes center-nose/nose-jaw	dist(12,18)/dist(18,31)
	eye center-face center/face width	eye center-face center/face width	dist(mid(13,11),12)/dist(19,23)

3.1.3. Dataset and Facial Attractiveness Rating by Humans

We used a collection of 80 frontal face image of Asian women with neutral expression. Half were randomly selected for training, and the remainder were used for testing. Thirteen human assessors (five men and eight women) were asked to score the attractiveness of each face on a seven-point scale (one being the least attractive and seven being the most attractive). There is subjectivity in each rater's scoring [24]. For example, a score of four could be considered a high score for one rater, and an average score for another rater, we applied both z-score normalization and linear scaling transformation to keep the score values of each human rater within the same range. First, given a score set S of a human rater, the following z-score normalization was applied.

$$z_i = \frac{s_i - \text{mean}(S)}{\text{std}(S)} \quad (2)$$

where s_i and z_i denote the i th original and normalized score values, respectively, and $\text{mean}(\cdot)$ and $\text{std}(\cdot)$ denote the mean and standard deviation of a score set S . Then, to scale a score range from one (lower bound) to seven (upper bound), the z-scores are transformed according to:

$$z'_i = l_b + \frac{z_i - \min(Z)}{\max(Z) - \min(Z)} \times (u_b - l_b) \quad (3)$$

where l_b and u_b denote the lower bound and upper bound of a target score range, and $\min(\cdot)$ and $\max(\cdot)$ denote the minimum and maximum values of a given score set, respectively. Then, the final attractiveness score of each image was calculated by averaging each score from all human referees. Figure 4 shows the distribution of the average scores of each face in the training set (40 images).

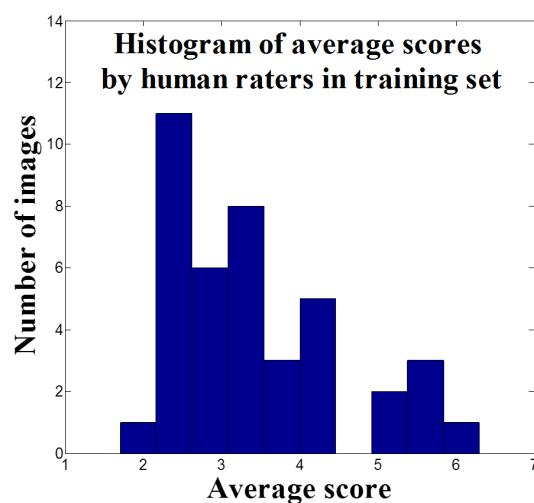


Figure 4. Histogram of average scores by human raters in the training set. The mean and standard deviation value of the scores are 3.42 and 1.12, respectively.

3.1.4. Predictor Construction

The predictors used in our framework were trained based on the feature sets (Section 3.1.2) and attractiveness ratings from humans (Section 3.1.3). In this study, we adopted three well-known regression-based predictors, support vector regression (SVR), K-nearest neighbor (K-NN) and artificial neural network (ANN). SVR transforms and analyzes data in a high-dimensional feature space mapped by a kernel function; accordingly, it is possible to arrive at an optimal linear function from a small amount of data. Here, we applied a Gaussian radial basis function (RBF) for kernel selection; parameters for training the RBF-SVR can be optimally selected during training and validation. K-NN regression is used to estimating facial attractiveness scores by analyzing a certain number K of the nearest feature vectors to the training samples. We used a weighted average of the K nearest neighbors, the weight of which was decided by the Euclidean distance of the K closest training samples. For ANN regression, we applied a multi-layer perceptron (MLP) composed of an input layer, hidden layers, and an output layer. Each layer has one or more neurons directionally linked with the neurons from the previous and next layers. A sigmoid function was applied as the activation function to compute the output of the hidden layer in each neuron. Due to the paucity of large training data, we used leave-one-out cross validation (LOOCV) to estimate the optimal parameter sets for each predictor. LOOCV involves using one training sample as a validation set and the remaining samples as the training set when building a target model. This process was repeated for each of the available training samples. Then, the optimal parameter sets allowing the lowest mean square error (MSE) were applied to build each predictor. The predictors then estimated attractiveness scores in a testing procedure. The parameter ranges considered in each score predictor are shown in Table 3.

Table 3. Parameter ranges of each score predictor.

Predictor	Parameter	Range
RBF-SVR	γ	0.1~10
	C	1~10
	ν	0.1~0.9
K-NN	K	1~30
ANN	No. hidden layer	1~2
	No. neurons	3~20

3.2. Testing Procedure

In testing, given a new image, RFSs were extracted by the same procedure as in the training step. Then, trained score predictors were applied to each RFS. Hence, for each RFS, three intermediate attractiveness scores were obtained. Even the score for each predictor can be utilized to estimate the attractiveness score, we applied a score fusion scheme to reduce the dependence on single score predictors and increase prediction performance. Score level fusion techniques are widely used in pattern recognition applications to enhance accuracy [45]. Among various kinds of score fusion schemes (average, product, maximum, minimum rule, etc.), average rule performed best in our case.

Table 4 shows the Pearson correlation value according to different fusion scenarios (different predictor fusion). As shown in Table 4, average score of three predictors has the best performance compared to the performance of the single or average score of only two predictors. Accordingly, for each RFS, the average score ($S^{RFS} = \text{avg}(S_{SVR}, S_{KNN}, S_{ANN})$) of the three predictors was calculated. Then the final attractiveness score (S^F) was calculated as follows:

$$S^F = \text{avg}(S^{RFS-1}, S^{RFS-2}, S^{RFS-3}, S^{RFS-4}) \quad (4)$$

Table 4. The performance (Pearson correlation value) comparison of different fusion scenarios.

Feature/Predictor	SVR	KNN	ANN	SVR + KNN	ANN + KNN	SVR + ANN	SVR + ANN + KNN
RFS-1	0.459	0.376	0.504	0.471	0.493	0.525	0.567
RFS-2	0.457	0.486	0.544	0.492	0.554	0.541	0.569
RFS-3	0.446	0.445	0.436	0.449	0.441	0.453	0.466
RFS-4	0.511	0.484	0.536	0.521	0.526	0.536	0.553

4. Experimental Results

4.1. Comparison with Human-Level Performance

The aim of this evaluation was to analyze the performance of each RFS with that of human assessors and to validate the performance improvement by our framework. As described in Section 3.1.3, 40 female faces were evaluated on a seven-point scale by 13 human assessors; additionally, SVR, KNN and ANN predictors were applied to each of the four RFSs and their average was the final attractiveness score. We therefore compared how human scoring correlates with machine scoring of attractiveness measures. Comparative performance was evaluated using Pearson correlation and MSE. As shown in Figure 5 and Table 5, each feature set has relatively high correlation with human raters in the test set.

In Figure 5, the x-axis represents scores from human raters on a seven-point scale, and the y-axis denotes corresponding prediction values from our framework. Neoclassical canons (RFS-1) and golden ratio values (RFS-2) have similar high correlation values. However, in the case of symmetry (RFS-3), the distribution of predicted values is more scattered, which corroborates studies that have shown that symmetry has less influence on facial attractiveness than other measures [12,20,21]. Score fusion for each RFS shows the highest correlation with human judgement. Since the factors that determine facial attractiveness are diverse and complicated, a single feature does not entirely estimate the attractiveness of a face, verifying that fusing complementary feature sets can improve attractiveness evaluation performance.

Also, we performed multiple regression analysis based on second order polynomial model to find which measure is the most predictive of human scores. As shown in Figure 5f, the regression curve for score fusion (blue curve) is the most similar with the ideal case (black dotted line). The second similar measure is RFS-2, and the less predictive ones are RFS-1 and RFS-3. As a result, fusion of four feature sets was found to be the most predictive of the human scores. After fitting data with our model, we have to evaluate the goodness of fit. To calculate the goodness of fit, we applied the well-known R-Square (R^2) statistic. R^2 measures how successful the fit is in explaining the variation of the data so

it represents the reliability of the predicted regression function. It is defined as the ratio of the sum of squares of the regression and the total sum of squares. It can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model. As shown in Table 5, the R^2 value of score fusion is the highest compared to other measure (four RFSs). Consequently, it was confirmed that score fusion is the most predictive of the human score among the attractiveness scores of each RFS.

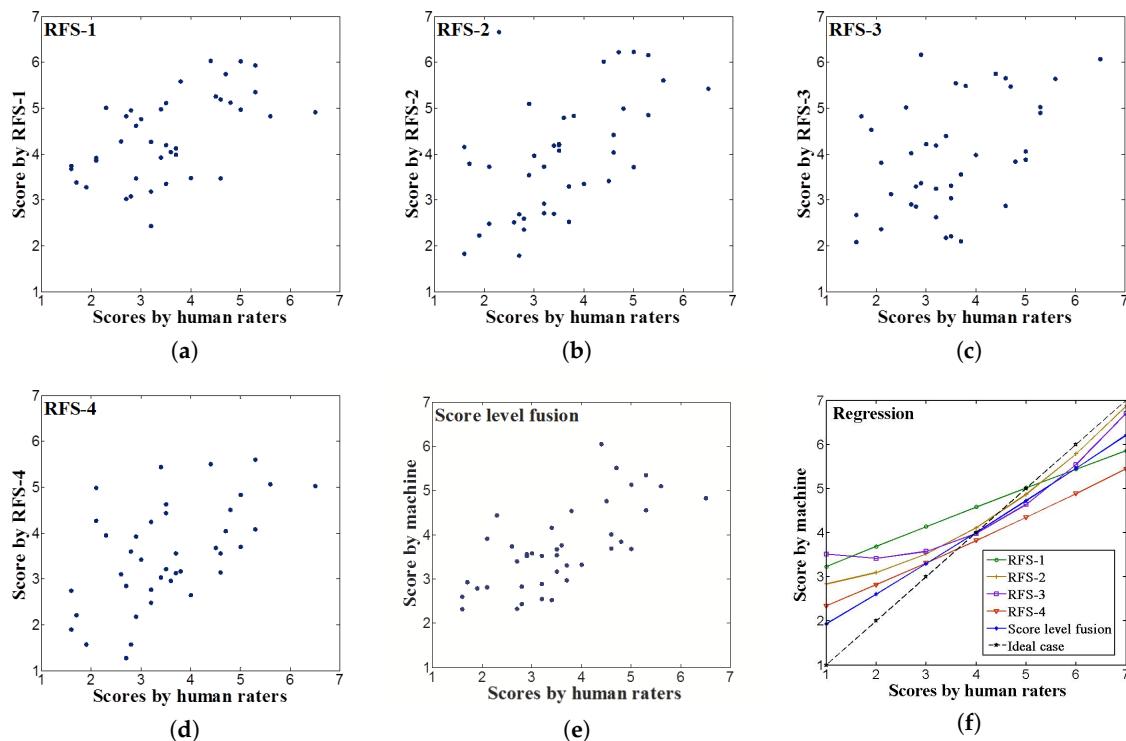


Figure 5. Correlation between predicted scores and average human scores in the test set (a–e) and regression curves of predictive scores and average human scores in the test set (f): (a) RFS-1; (b) RFS-2; (c) RFS-3; (d) RFS-4; (e) score level fusion; (f) regression curves.

Table 5. Correlation, MSE and R^2 between the predicted and human scores.

Feature Set	Pearson Correlation	MSE	R^2
RFS-1	0.567	0.206	0.282
RFS-2	0.569	0.193	0.301
RFS-3	0.466	0.220	0.197
RFS-4	0.553	0.170	0.241
Score fusion	0.689	0.151	0.449

4.2. Applications

Our framework can be applied to many areas, such as beauty ranking and plastic surgery. For example, Figure 6 demonstrates the predicted attractiveness scores of two female Korean celebrities. The two celebrities have relatively higher attractiveness scores than the average in our test set (3.69). Figure 7 shows the validation of a facial beauty enhancement method [27] using the proposed framework. An increase in beauty score is observed from the original to the beauty-enhanced face image. Hence, it is possible to evaluate surgery results, or to recommend methods of attractiveness enhancement based on our framework.

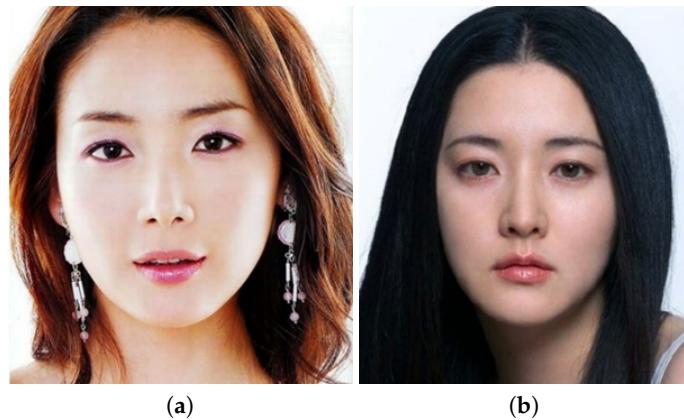


Figure 6. Examples of predicted attractiveness score. Two Korean female celebrities, (a,b), obtain scores of 5.1 and 5.34, respectively. Note that the average and standard deviation values of our test set are 3.69 and 0.69, respectively.

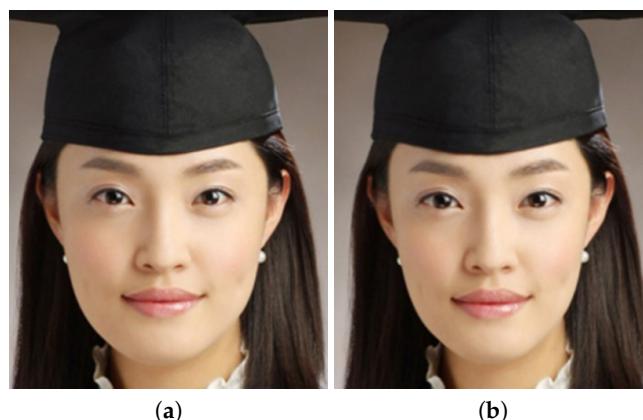


Figure 7. Examples of the validation of a facial beauty enhancement system [27]. The facial attractiveness score increases from 2.82 (original face (a)) to 4.44 (beauty enhanced face (b)) after applying the facial beauty enhancement system.

5. Conclusions

In this study, we have proposed a novel framework for assessing facial attractiveness. The proposed framework utilizes four types of ratio feature sets derived from universal standards of facial beauty. To enhance the system's performance to be comparable to human rating, three types of regression-based predictors were incorporated to estimate an attractiveness score, and score level fusion was performed. Experimental results showed that the attractiveness score obtained by score fusion better correlates with those of human assessors than scores from other predictors, indicating that a fusion of multiple facial proportion features, rather than a single feature, can provide better performance for automatic facial beauty evaluation and enhancement. In addition, our result showed that symmetry performed more poorly than other proportion-based features, supporting the conclusion of other studies.

Our proposed method has shown that simple proportion-based facial attractiveness measurements, such as symmetry, the golden ratio, and neoclassical canon, are associated with human judgement. Combining complementary proportion features can improve prediction performance and it is the most predictive of the human scores. Moreover, the modularity of the proposed framework enables any features or predictors to be integrated into automatic facial attractiveness measurements. Therefore, the framework can be utilized in various attractiveness evaluation and enhancement applications. Our future work consists of consolidating the framework by increasing the face dataset

with diverse human referees; adopting more facial shape features, such as averageness and femininity; considering facial appearance features such as colors, tones, and texture; finding the most predictive features and modeling method compatible with human scoring; analyzing face attractiveness in various races and genders; and extending the framework to measure the facial impressions (trustworthiness, dominance, etc.) in social dimensions.

Acknowledgments: This work was supported by a KIST Institutional Program [Project No. 27190 and No. 2V05260].

Author Contributions: I.K. and H.C. conceived and designed the experiments; Y.H. performed the experiments; Y.H. and H.C. analyzed the data; G.P.N. and J.C. contributed analysis tools; Y.H. wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rhodes, G. The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* **2006**, *57*, 199–226.
2. Said, C.P.; Todorov, A. A statistical model of facial attractiveness. *Psychol. Sci.* **2011**, *22*, 1183–1190.
3. Adamson, P.A.; Galli, S.K.D. Modern concepts of beauty. *Plast. Surg. Nurs.* **2009**, *29*, 5–9.
4. Perrett, D.I.; May, K.A.; Yoshikawa, S. Facial shape and judgements of female attractiveness. *Nature* **1994**, *368*, 239–242.
5. Winston, J.S.; O'Doherty, J.; Kilner, J.M.; Perrett, D.I.; Dolan, R.J. Brain systems for assessing facial attractiveness. *Neuropsychologia* **2007**, *45*, 195–206.
6. Rubenstein, A.J.; Kalakanis, L.; Langlois, J.H. Infant preferences for attractive faces: A cognitive explanation. *Dev. Psychol.* **1999**, *35*, 848–855.
7. Langlois, J.H.; Roggman, L.A. Attractive faces are only average. *Psychol. Sci.* **1990**, *1*, 115–121.
8. Rhodes, G.; Proffitt, F.; Grady, J.M.; Sumich, A. Facial symmetry and the perception of beauty. *Psychon. Bull. Rev.* **1998**, *5*, 659–669.
9. Fink, B.; Grammer, K.; Thornhill, R. Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. *J. Comp. Psychol.* **2001**, *115*, 92–99.
10. Jones, A.L.; Russell, R.; Ward, R. Cosmetics Alter Biologically-Based Factors of Beauty: Evidence from Facial Contrast. *Evol. Psychol.* **2015**, *13*, 210–229.
11. Kagian, A.; Dror, G.; Leyvand, T.; Cohen-Or, D.; Ruppin, E. A humanlike predictor of facial attractiveness. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3 December 2007; pp. 649–656.
12. Jones, A.L. The influence of shape and colour cue classes on facial health perception. *Evol. Hum. Behav.* **2017**, in press.
13. Farkas, L.G.; Hreczko, T.A.; Kolar, J.C.; Munro, I.R. Vertical and horizontal proportions of the face in young adult north american caucasians: A revision of neoclassical canons. *Plast. Reconstr. Surg.* **1985**, *75*, 328–337.
14. Narain, D.L. The Golden Ratio. Available online: <http://cuip.uchicago.edu/~dlnarain/golden/> (accessed on 5 April 2017).
15. Jefferson, Y. Facial beauty: Establishing a universal standard. *Int. J. Orthod.* **2004**, *15*, 9–22.
16. Galton, F. Composite portraits, made by combining those of many different persons into a single resultant figure. *J. Anthropol. Inst. Greatt. Br. Irel.* **1879**, *8*, 132–144.
17. Thomas, R.A.; Michael, R.C. Averaged faces are attractive, but very attractive faces are not average. *Psychol. Sci.* **1991**, *2*, 123–125.
18. Grammer, K.; Thornhill, R. Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *J. Comp. Psychol.* **1994**, *108*, 233–242.
19. Perrett, D.I.; Lee, K.J.; Penton-Voak, I.; Rowland, D.; Yoshikawa, S.; Burt, D.M.; Henzi, S.P.; Castlesm, D.L.; Akamatsu, S. Effects of sexual dimorphism on facial attractiveness. *Nature* **1998**, *394*, 884–887.
20. Scheib, J.E.; Gangestad, S.W.; Thornhill, R. Facial attractiveness, symmetry and cues of good genes. *Proc. Biol. Sci.* **1999**, *266*, 1913–1917.
21. Bronstad, P.M.; Russell, R. Computational models of facial attractiveness judgments. *Perception* **2008**, *37*, 126–142.

22. Jones, B.C.; Little, A.C.; Burt, D.M.; Perrett, D.I. When facial attractiveness is only skin deep. *Perception* **2004**, *33*, 569–576.
23. Aarabi, P.; Hughes, D.; Mohajer, K.; Emami, M. The automatic measurement of facial beauty. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Tucson, AZ, USA, 7–10 October 2001; Volume 4, pp. 2644–2647.
24. Gunes, H.; Piccardi, M. Assessing facial beauty through proportion analysis by image processing and supervised learning. *Int. J. Hum.-Comput. Stud.* **2006**, *64*, 1184–1199.
25. Eisenthal, Y.; Dror, G.; Ruppin, E. Facial attractiveness: Beauty and the machine. *Neural Comput.* **2006**, *18*, 119–142.
26. Schmid, K.; Marx, D.; Samal, A. Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratio. *Pattern Recogn.* **2008**, *41*, 2710–2717.
27. Leyvand, T.; Cohen-Or, D.; Dror, G.; Lischinski, D. Data-driven enhancement of facial attractiveness. In Proceedings of the ACM SIGGRAPH, Los Angeles, CA, USA, 11–15 August 2008; pp. 38.1–38.9.
28. O’Toole, A.J.; Price, T.; Vetter, T.; Bartlett, J.C.; Blanz, V. 3D shape and 2D surface textures of human faces: The role of “averages” in attractiveness and age. *Image. Vis. Comput.* **1999**, *18*, 9–19.
29. Fink, B.; Grammer, K.; Matts, P.J. Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evol. Hum. Behav.* **2006**, *27*, 433–442.
30. Jang, K.S.; Bayome, M.; Park, J.H.; Park, K.H.; Moon, H.B.; Kook, Y.A. A three-dimensional photogrammetric analysis of the facial esthetics of the Miss Korea pageant contestants. *Korean J. Orthod.* **2017**, *47*, 87–99.
31. Vezzetti, E.; Marcolin, F.; Tornincasa, S.; Ulrich, L.; Dagnes, N. 3D geometry-based automatic landmark localization in presence of facial occlusions. *Multimed. Tools Appl.* **2017**, in press.
32. Marcolin, F.; Vezzetti, E. Novel descriptors for geometrical 3D face analysis. *Multimed. Tools Appl.* **2017**, *76*, 13805–13834.
33. Liao, Q.; Jin, X.; Zeng, W. Enhancing the Symmetry and Proportion of 3D Face Geometry. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1704–1716.
34. Cootes, T.; Taylor, C.; Cooper, D.; Graham, J. Active shape models—Their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59.
35. Milborrow, S.; Nicolls, F. Locating facial features with an extended active shape model. In Proceedings of the ECCV, Marseille, France, 12–18 October 2008; pp. 504–513.
36. Torsello, F.; Mirigliani, L.; D’Alessio, R.; Deli, R. Do the neoclassical canons still describe the beauty of faces? An anthropometric study on 50 Caucasian models. *Prog. Orthod.* **2010**, *11*, 13–19.
37. Mommaerts, M.Y.; Moerenhout, B.A. Ideal proportions in full face front view, contemporary versus antique. *J. Craniomaxillofac. Surg.* **2011**, *39*, 107–110.
38. Naini, F.B.; Moss, J.P.; Gill, D.S. The enigma of facial beauty: Esthetics, proportions, deformity, and controversy. *Am. J. Orthod. Dentofacial. Orthop.* **2006**, *130*, 277–282.
39. Pancherz, H.; Knapp, V.; Erbe, C.; Heiss, A.M. Divine proportions in attractive and nonattractive faces. *World J. Orthod.* **2010**, *11*, 27–36.
40. Magne, P.; Gallucci, G.O.; Belser, U.C. Anatomic crown width/length ratios of unworn and worn maxillary teeth in white subjects. *J. Prosthet. Dent.* **2003**, *89*, 453–461.
41. Ong, E.; Brown, R.A.; Richmond, S. Peer assessment of dental attractiveness. *Am. J. Orthod. Dentofacial. Orthop.* **2006**, *130*, 163–169.
42. Cunningham, M.R. Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *J. Personal. Soc. Psychol.* **1986**, *50*, 925–935.
43. Johnston, V.S.; Franklin, M. Is beauty in the eye of the beholder? *Ethol. Sociobiol.* **1993**, *14*, 183–199.
44. Etcoff, N.L. Beauty and the beholder. *Nature* **1994**, *368*, 186–187.
45. Jain, A.; Nandakumar, K.; Ross, A. Score normalization in multimodal biometric systems. *Pattern Recogn.* **2005**, *38*, 2270–2285.

