Automated Classification of Female Facial Beauty Using Learning Algorithms

Hatice Gunes, Massimo Piccardi, Tony Jan
Computer Vision Group –Faculty of Information Technology
University of Technology, Sydney (UTS)
PO Box 123 Broadway 2007 NSW – Australia
e-mail: {haticeg,massimo,jant}@it.uts.edu.au

Abstract

Among artists and psychologists, there exists a widespread belief that facial attractiveness can be measured by using mathematical ratios of facial features. Accordingly, in this paper we present an automated feature-based measuring system attempting to measure female facial beauty in a quantitative and repeatable way. The system is based on feature measurements made with image analysis algorithms and automated classification based on supervised learning. The system proves able to achieve good accuracy, substantially confirming the objective measurability of female facial beauty.

Keywords: facial beauty classification, golden proportion, facial features, supervised learning

1. Introduction

It has long been believed that the concept of facial beauty is subjective and varies by race, culture or era. However, mathematics and medical science state that there is a timeless aesthetic ideal for facial beauty and our perception of physical beauty is based mostly on our physical proportions' closeness to phi (the Golden Ratio, namely 1.61803 39887 49894 84820 [1,2]). Human beauty has been studied for years in practices of oral and maxillo-facial surgery [3]. Recent scientific studies revealed that the concept of a beautiful face is formed at an age of 2 months. They state that the concepts of a "beautiful face" are not learned but are "hard-wired" into our brain from the very beginning [1,2]. Moreover, the cross-cultural surveys on facial beauty show that all groups have similar perception of facial beauty, and the concept of beauty can be defined for all races, cultures and eras with right facial proportions [3].

The concept of beauty is obvious to the observer but is difficult to quantify. However, the science of measuring beauty existed for centuries. Through history and across different cultures, investigators in psychology, arts and image analysis have extensively studied quantifying the facial beauty. There is much evidence of studies for measuring facial beauty [4-15], the most famous of these being the Golden Proportions (based on the Golden Ratio) [14] and the Facial Thirds [15].

Measuring female facial beauty from a still image is not a trivial task since there doesn't exist a generally agreed measurement approach. The group from Marquardt Aesthetic Imaging Inc. has developed a method that measures facial beauty using the Golden Ratio [3]. Major limitations of this method are the lack of objective confirmation and the fact that it requires many manual measurements. Instead, a fully automated procedure could be much faster and provide more objective and repeatable results. Aarabi and Hughes in [16] were able to propose an automated procedure. However, their beauty grading system was not explicit and this might limit the consensus about the proposed approach. In a previous paper from one of the authors of this paper, an automated procedure was proposed based on image analysis and a rule-based classification system [17]. The classification rules were elicited by hand and classification compared with that from a pool of human referees.

In this paper, we propose instead a classification system based on supervised learning able to automatically learn the classification rules. Experiments were performed with data sets from different ethnicity to explore the dependence of the classification rules on ethnical characters. The experimental results prove that the automated classification of facial beauty is possible and is in good accordance with human classification. Moreover, the ancient intuition on Golden Proportions seems substantially confirmed.

The rest of the paper is organized as follows: Section 2 describes the general rules for measuring facial beauty and presents the proposed approach. Section 3 describes the face detection process and the feature extraction operators. Section 4 describes the classification process. Section 5 presents and discusses the experimental results, and, finally, Section 6 presents the conclusions and addresses future work.

2. Methods Used for Measuring Beauty

In this section we provide information on the general rules applied and give an overview of the method.

2.1. General Rules Applied

The *Golden Proportions* and *Facial Thirds* are ratios derived from specific facial features [3,15,18].

The *Golden Ratio or Proportion* is approximately the ratio of 1 to 0.618 or the ratio of 1.618 to 1 as shown in Figure 1. For a perfect, vertically aligned face, all the proportions stated in Table 1 must fit the Golden Proportion 1.618.

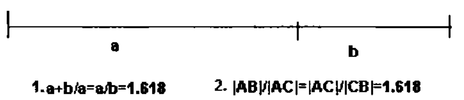
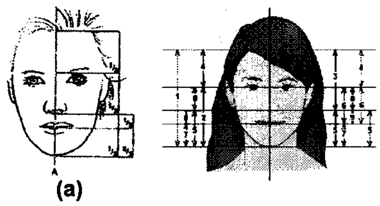


Figure 1. Golden proportion.

Table 1: Ratios used for golden proportion comparison.

2:4 Ratio of vertical distance between *Eyes* and *Chin* to vertical distance between *Forehead* and *Eyes*
3:5 Ratio of vertical distance between *Forehead* and *Nose* to vertical distance between *Nose* and *Chin*
6:7 Ratio of vertical distance between *Eyes* and *Lips* to vertical distance between *Lips* and *Chin*
5:8 Ratio of vertical distance between *Nose* and *Chin* to vertical distance between *Eyes* and *Nose*
8:9 Ratio of vertical distance between *Eyes* and *Nose* to vertical distance between *Nose* and *Lips*
7:9 Ratio of vertical distance between *Lips* and *Chin* to vertical distance between *Nose* and *Lips*

Facial Thirds state that a well-proportioned face may be divided into roughly equal thirds by drawing horizontal lines through the forehead hairline, the brow, the base of the nose, and the edge of the chin [3,15,18].



(a)

Figure 2. a) Facial thirds; b) Template image with golden proportions.

Table 2: Ratios used for facial thirds measurements.

Segment	Measurement Type
Distance between forehead and eyebrows	"Facial Thirds" state that all of the three segments should be equal
Distance between eyebrows and nose	
Distance between nose and chin	
Distance between lips and chin	The distance between lips and chin should be the double than the distance between nose and lips
Distance between nose and lips	

According to these rules, a face is more attractive as it approaches the proportions shown in Table 1 and Table 2. Facial plastic surgeons have long been using these ratios as a guide for their work [3, 13].

2.2. Overview of the Applied Method

In our research, the first step in measuring facial beauty is to extract the features of a female face. This face feature extraction task is accomplished in six phases: face localization by skin region detection, eye localization, pupil localization, eyebrow localization, lip localization by color segmentation, nose localization and finally, chin localization. The last phase of our method involves estimating the proportions of the face from the features extracted and measuring the beauty of the face. This estimation is based on the two facial beauty estimation methods described in the previous sub-section.

3. Feature Extraction and Analysis

Accurate feature extraction is fundamental for reliable and precise measuring of facial beauty. A vast literature covers techniques for facial feature extraction ([19-22] are just a few examples of very recent proposals). In this work, we chose to use simple algorithms able to be adequately accurate and computationally efficient.

Detection of Facial Region

It is important to locate the facial region in order to remove irrelevant picture information. In order to obtain the exact location of the face, a non-skin region filter is applied [17,23].

Eye Localization

After defining the location of the facial region, detecting the eye region is fast and simple. First, the vertical histogram of the skin-colored regions is computed [23]. Then, the rows containing the eyes are located in correspondence of a histogram local minimum in the high histogram part. Finally, the horizontal histogram for these rows is computed, and eyes located as the two local minimums.

Eyebrow Localization

After detecting the pair of possible eyes which satisfies the geometrical constraints imposed by the face, it is easier to localize the eyebrows. Eyebrows are expected to be located in the upper part of the face and are the first non-skin components on the facial region down the forehead.

Lip Localization

We choose to locate the lips after having located the eyes because their horizontal position is in-between those of the eyes; lips can be easily discriminated from skin based on color.

Nose Localization

Nose anatomically is located between the eyes and the mouth. Searching for the nose is relatively easy due to the well-confined search space limited by lips and pupils.

Chin Localization

The chin detection process takes place after lip detection, as the chin anatomically is located between the lips and the neck. The search space is arranged according to the lip line and the horizontal lower limit of the facial region.

4. Facial Beauty Classification

Our main assumption so far is that the features extracted from images will be able to support accurate classification of facial beauty. In the following, we describe the approach for generating automatic classification rules.

In our previous work [17], we manually generated rules based on the Golden Proportions, Facial Thirds and other measurements, according to a scoring scheme used by humans. Since this process was too laborious and demanding in terms of human effort, in this work we approached classification with automatic classification rules generation.

The automatic classification rules provided by a classifier generator should meet the following requirements:

- the rules must provide accurate classification;
- the classification rules should be possibly interpretable and shareable by human experts;
- the rule system must be simple enough to avoid unnecessary large computations.

To cope with these requirements, we experimented the supervised classifier C4.5 [24,25]. C4.5 generates decision trees or rule sets based on the notion of entropy. Supervised classification requires the user to pre-classify a set of samples and train the classifier based on these samples. The training set must be significant, representing a complete set of possible cases.

Features extracted from each face were stored as tuples of attributes characterizing the face, and used for defining the classification rules. The actual feature set contains feature measurements, inter-feature distances, ratios from golden proportion and facial thirds, and two measurements empirically assessing the overall deviation from golden proportions and facial thirds (GP_Total_Difference and FT_Total_Difference, respectively). The full feature set is shown in Fig. 3.

1	Face Length	13	Ratio PL LC
2	Face Width	14	Ratio NC PN
3	Pupil to Chin	15	Ratio PN NL
4	Forehead to Nose	16	Ratio LC NL
5	Forehead to Pupils	17	Mean Ratio
6	Nose to Chin	18	GP Total Difference
7	Pupils to Lips	19	Forehead to Eyebrow
8	Lips to Chin	20	Eyebrow to Nose
9	Pupils to Nose	21	Ratio FrhEyebrow M
10	Nose to Lips	22	Ratio Eyebrow Nose M
11	Ratio PC FP	23	Ratio Nose Chin M
12	Ratio FN NC	24	FT Total Difference

Figure 3. The feature set for beauty classification.

The feature set in Fig. 3 gives rise to a 24-dimension classification space which must be partitioned into classes by C4.5. C4.5's decision trees split this classification space into hyperrectangles, in the sense that each tree node states a comparison between the value of one of the features at a time against a threshold. A leaf is a terminal node corresponding with a hyper rectangle of minimal size. A class is defined by its set of leaves [26].

5. Results and Analysis

For our experiments, we pre-classified a set of 71 mixed-race faces [26] into 3 classes, namely "beautiful" ("b", or Type 1), "average" ("a", or Type 2), and unattractive ("u", or Type 3}; each example was pre-classified by voting of 5 individuals. Later, we restricted the set to 45 Turkish female faces, in order to explore potential differences in the classification due to ethnicity. For tree generation, we used 10-fold cross validation [25].

Preliminary experiments performed with the complete feature set led to very complex classification trees of difficult interpretation and convinced us to use a sub-set of the initial feature set containing only the Golden Proportion features (features 1, 3-18). Fig. 4 shows the classification tree for the mixed-race data set. Tests have been numbered in order to ease reference. The root node (referred with (1) and (1') in Fig. 4) tests feature GP_Total_Difference, which is the sum of the total difference between each ratio's actual value and the Golden Proportion; in the ideal case, this feature should be 0. Test (1/1') actually splits the tree into two sub-trees: in that with GP_Total_Difference <= 1.16, there are instances from the "beautiful" and "average" classes; in the other, instances from "average" and "unattractive". This confirms that this test is in accordance with the Golden Proportion rule.

The tree contains other four tests, one based on GP_Total_Difference (2/2') and the other three on single ratios; each single ratio should be as close as possible to 1.618. Actually, tests (3/3') and (5/5') are in accordance with the Golden Proportion, while (2/2') and (4/4') are not. This means that the Golden Proportion proves a maximal rule, only partially confirmed by this experiment. However, the perfectly beautiful instance according to Golden Proportions (GP_Total_Difference = 0 and all single ratios = 1.618) will be correctly classified as "beautiful" under leaf (3).

```
        J48 pruned tree
        ------------------

(1)     GP_Total_Difference <= 1.16
(2)     |   GP_Total_Difference <= 0.7228
(3)     |   |   Ratio_FN_NC <= 1.757: b
(3')    |   |   Ratio_FN_NC > 1.757: a
(2')    |   GP_Total_Difference > 0.7228
(4)     |   |   Ratio_PC_FP <= 1.407: b
(4')    |   |   Ratio_PC_FP > 1.407: a
(1')    GP_Total_Difference > 1.16
(5)     |   Ratio_PL_LC <= 1.294: u
(5')    |   Ratio_PL_LC > 1.294: a

        Number of Leaves  :     6

        Size of the tree :      11
```

Figure 4. Decision tree for the mixed-race data set.

```
     J48 pruned tree
     ------------------

(1)
     GP_Total_Difference <= 1.097
     |   Ratio_LC_NL <= 1.789
     |   |   Ratio_LC_NL <= 1.579: a
     |   |   Ratio_LC_NL > 1.579
     |   |   |   Ratio_FN_NC <= 1.642: b
(b)  |   |   |   Ratio_FN_NC > 1.642
     |   |   |   |   Ratio_LC_NL <= 1.737: b
     |   |   |   |   Ratio_LC_NL > 1.737: a
     |   Ratio_LC_NL > 1.789: a (4.0)
(1')
     GP_Total_Difference > 1.097
     |   Ratio_PL_LC <= 1.647
     |   |   GP_Total_Difference <= 1.5: a
     |   |   GP_Total_Difference > 1.5: u
     |   Ratio_PL_LC > 1.647: a

     Number of Leaves  :     8

     Size of the tree :      15
```

Figure 5. Decision tree for the Turkish data set.

Fig. 5 shows the classification tree for the Turkish data set. Even if this tree is not identical to that for the mixed-race data set, the tree structure and semantic are very similar. The node test (1/1') is again on feature GP_Total_Difference and in accordance with the Golden Proportion rule. The tree contains other six tests, in substantial accordance with the Golden Proportion. Once again, the perfectly beautiful instance according to Golden Proportion is correctly classified as "beautiful", under leaf (b).

It is important to note that, while the Golden Proportion states a rule for assessing perfect beauty, it does not provide any rule for assessing partial beauty (can the beauty be considered to degrade *linearly* as the ratio value moves away from 1.618? Higher values are worse than lower values, or vice versa?). This problem is particularly relevant in plastic surgery, where it is evident that facial features cannot be arbitrarily modified to reach exactly Golden Proportions. In this context, how can two feasible modifications, none of them exactly providing Golden Proportions, be compared in terms of achievable beauty improvement? The classifier generated in this work, able to classify any combination of facial feature values, provides convincing answers to these questions. Fig. 6 shows an example of faces differently classified by our system.

Figure 6. Example of faces differently classified by the system.

Table 3: Comparison of human and system classification (mixed-race data set).

S.C. H.C.	Type 1	Type 2	Type 3
Type 1	6	2	1
Type 2	1	30	11
Type 3	1	3	16

H.C: Human Classification->True
S.C: System Classification->Correct/ Incorrect

Table 3 reports a classification matrix comparing system classification against that of human referees. The accuracy achieved is high (the major diagonal of Table 3 reports the correct results), although there are also numerous mismatches. In particular, the system under-classified as Type 3 several faces classified as Type 2 by the human referees. However, it is important to note that human classification cannot be classified as "exact" *tout-court*, since human classification by different referees is not always identical. This difference in the human referees' evaluation might be enough to question the initial hypothesis that beauty can be objectively measured. However, our experiments gave reasonable evidence that measurable facial features and supervised learning allow to build a classification system generally consistent with the average human judgement.

6. Conclusions and Future Work

In this paper, we have proposed an automated system for female facial beauty classification. The system can be considered an evolution of a previous system from one of these authors, where the classification rules were elicited manually [17].

This paper has given substantial evidence to the fact that facial beauty can be objectively classified based on a set of measurements from facial features. The classification provided by the system is in good accordance with the average classification from a pool of human referees. However, it is likely that other features influence the classification of facial beauty (color and shape of features such as eyes, lips, nose, etc; smiling vs. angry attitude, and others).

The experimental results are in good accordance with the ancient intuition of Golden Proportions. In addition, experiments performed on a mixed-race and single ethnicity (Turkish) face sets did not report relevant differences, therefore confirming the idea that the concept of beauty might be universal.

In the near future, we aim to extend the experimental results to larger and more varied databases such as FERET [27]. Our future work will be aimed at training our system to measure female facial beauty into more refined classes, in order to improve the capability of discriminating between different beauty grades. Moreover, we want to extend the features set to cover other influencing features such as eyes and lips' color and shape.

A major aim of the analysis done in this paper is to providing suggestions on how to improve female facial beauty, such as surgery, makeup or hairstyle improvements. Eventually, extending the classification to male facial beauty will be considered.

References

[1] LARRABEE, W., "Facial Beauty: Myth or Reality?", ARCHIVES OF OTOLARYNGOLOGY -Head & Neck Surgery, Vol 123:571-572, 1997.

[2] YELLIN, S. "Aesthetics for the Next Millenium", FACIAL PLASTIC SURGERY MONOGRAPHS, Vol 13(4):231-239, 1997.

[3] Marquardt Beauty Analysis, http://www.beautyanalysis.com/index2_mba.htm.

[4] BELL , A. (1997), The Definition of Beauty, NATURE, October/November issue.

[5] PERRETT, D.I., MAY, K.A. & YOSHIKAWA, S. (1994). Facial shape and judgments of female attractiveness. NATURE, 239-242.

[6] LANGLOIS, J.H. & ROGGMAN, L.A. (1990). Attractive faces are only average. Psychological Science, 1, 115-121.

[7] CUNNINGHAM , M.R., ROBERTS, A.R, BARBEE, A.P., DRUEN, P.B et al. (1995). "Their ideas of beauty are, on the whole, the same as ours", Journal of Personality & Social Psychology, 68, 261-279.

[8] JEFFERSON , Y. (1993). Facial aesthetics--presentation of an ideal face. Journal of General Orthodontics, 4, 18-23.

[9] MEALEY , L., BRIDGSTOCK, R, TOWNSEND, G.C. (1999). Symmetry and perceived facial attractiveness: A monozygotic co-twin comparison.

[10] LANDAU, T. About Faces, Bantam Doubleday Dell Publishing Group, Inc. 1989. New York.

[11] MICHIELS, G. & SATHER A.H. (1994). Determinants of facial attractiveness in a sample of white women. International Journal of Adult Orthodontics & Orthognathic Surgery, 9, 95-103.

[12] DAIBO, IKUO, Suggestion from comparison research of facial beauty AASP 1999 (third) Conference in Taipei August 4 1999, Hokusei Gakuen University, Sapporo, Japan.

[13] PARRIS , CYNARA, The Bold and the Beautiful (According to Plastic Surgeons), Tyler Street Christian Academy, Dallas, Texas, Jack Robinson, Jr., Ph.D.
[14] BBC Science - the human FACE, http://www.bbc.co.uk/science/humanbody/humanface/beauty_golden_mean.shtml.
[15] FARKAS , LESLIE G. et al., 1985,Vertical and horizontal proportions of the face in young adult North American Caucasians, Plastic and Reconstructive Surgery 75(3): 328-38.
[16] AARABI, P., HUGHES, D., The automatic measurement of facial beauty, 2001 IEEE International Conference on Systems, Man, and Cybernetics, Page(s): 2644 -2647 vol.4, 2001.
[17] GUNES, H., KARSLIGIL, M.Y., Measuring Female Facial Beauty by Calculating the Proportions of the Face, afp Proc. of ISCIS XVII Seventeenth Int. Symp. on Computer and Information Sciences, Oct. 2002, Orlando, Florida, USA.
[18] RICKETTS, M.D., 1982. Divine proportions in facial aesthetics. Clinics in Plastic Surgery Vol. 9, No. 4.
[19] REIN-LIEN HSU, ABDEL-MOTTALEB, M.; JAIN, A.K. (2002), Face detection in color images, IEEE Transactions on Pattern Analysis and Machine Intelligence, on , Volume: 24 Issue: 5 , 2002 , pp. 696 –706.
[20] KAPOOR, A., PICARD, R.W. (2002) , Real-time, fully automatic upper facial feature tracking, Proc. of Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2002, pp. 10 – 15.
[21] YEN, G.G., Nithianandan, (2002), Facial feature extraction using genetic algorithm, Proc. of the 2002 Congress on Evolutionary Computation, 2002, vol. 2, pp. 1895 – 1900.
[22] FERIS, R.S., GEMMELL, J., TOYAMA, K., KRUGER, V. (2002), Hierarchical wavelet networks for facial feature localization, Proc. of Fifth IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2002, pp. 125 – 130.
[23] CHENG, J., (2001), Boston University, CS585 Image and Video Computing.
[24] R. S. MICHALSKI , J. G. CARBONELL, and T. M. MITCHELL, editors, Machine Learning - An Artificial Intelligence Approach. Springer-Verlag, Berlin, 1984.
[25] J. R. QUINLAN, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, California, 1993
[26] H. J. HAMILTON, course notes, Univ. of Regina, Canada http://www.cs.uregina.ca/%7Ehamilton/courses/831/notes/ml/dtrees/4_dtrees1.html.
[27] The Facial Recognition Technology (FERET) Database, http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
[28] A.M. MARTINEZ and R. BENAVENTE. *The AR Face Database*. CVC Technical Report #24, June 1998.