



A computational approach to body mass index prediction from face images[☆]

Lingyun Wen, Guodong Guo^{*}

Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, 26506, United States



ARTICLE INFO

Article history:

Received 11 June 2012

Received in revised form 9 February 2013

Accepted 24 March 2013

Keywords:

Body mass index (BMI)

Facial features

BMI prediction

Machine vision

Large database

ABSTRACT

Human faces encode plenty of useful information. Recent studies in psychology and human perception have found that facial features have relations to human weight or body mass index (BMI). These studies focus on finding the correlations between facial features and the BMI. Motivated by the recent psychology studies, we develop a computational method to predict the BMI from face images automatically. We formulate the BMI prediction from facial features as a machine vision problem, and evaluate our approach on a large database with more than 14,500 face images. A promising result has been obtained, which demonstrates the feasibility of developing a computational system for BMI prediction from face images at a large scale.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Human faces contain a number of cues, e.g., identity, emotional expression, age, gender, ethnicity, attractiveness, personality traits, and so on. Decoding facial cues has been the subject of speculation for centuries [1]. It has attracted much attention from psychologists, sociologists, and computer scientists for decoding the useful information from faces.

In this paper, we want to “decode” the information of body mass index (BMI) from single face images. BMI is a measure of body fat based on body height and weight for an individual. Gallagher et al. [2] showed that BMI is representative of body fatness.

BMI is also an important visual characteristic to describe a person. BMI is a widely used measure of adiposity in humans, especially for the overweight issue. Andrew et al. [3] revealed that BMI is associated with the risk of some common adult cancers.

To compute an individual's BMI, the traditional method is usually to measure both the body weight and height. Thus some tools are required, such as a ruler and a scale.

Given an individual's height and weight, the calculation of body mass index (BMI) is given by [4]:

$$\text{BMI} = \begin{cases} \frac{\text{weight}(\text{kg})}{\text{height}(\text{m})^2} \\ \text{or} \\ \frac{\text{weight}(\text{lb}) \times 703}{\text{height}(\text{in})^2} \end{cases}$$

depending on the units to use, e.g., meter (m) for height and kilogram (kg) for weight, or inch (in) for height and pound (lb) for weight. According to the values of BMI, people are divided into four categories: underweight, normal, overweight, and obese [5]. The BMI categories and the range of values in each category are shown in Table 1. Also, Fig. 1 shows some face examples in the different categories according to their BMI values.

1.1. Related work

Recent studies in psychology and human perception [6–8] have found that facial features have relations to human weight or BMI. These studies focus on finding the related, specific features on faces, and measuring the correlations between facial features and the BMI.

Coetzee et al. [6] showed that facial adiposity, or the perception of weight in the face, can predict perceived health and attractiveness. They recruited 84 Caucasian participants (43 females and 41 males) to capture face photos. The weight, height, and other information, e.g., blood pressure, of all participants were recorded. Then they recruited another four groups of people to rate the facial images manually. They showed that the rated facial adiposity is related to BMI.

[☆] This paper has been recommended for acceptance by Lijun Yin.

^{*} Corresponding author. Tel.: +1 304 293 9143.

E-mail address: guodong.guo@mail.wvu.edu (G. Guo).

Table 1

BMI categories. People are divided into four categories based on BMI values: underweight, normal, overweight and obese.

Category	BMI range
Underweight	<18.5
Normal	18.5–24.9
Overweight	25.0–29.9
Obese	>30

Coetzee et al. [7] studied three facial features: width to height ratio, perimeter to area ratio, and cheek to jaw width ratio. They captured face photos of 95 Caucasian and 99 African participants. Each facial image was manually delineated by defining 179 feature points and aligned according to interpupillary distance using a computer software. Pearson's correlations were computed. They showed that the three facial features were significantly related to BMI in males, while only width to height ratio and cheek to jaw ratio were significantly related to BMI in females.

Pham et al. [8] studied correlations between seven facial features and BMI, in young and elderly people in Korean. They recruited 911 participants in two age groups: twenties and sixties. A well-trained operator was asked to manually label the facial features. Then the Pearson's correlation coefficients were calculated to assess the association of facial features to BMI.

1.2. Our approach

In addition to the evidence from psychology and human perception [6–8] on the relation between facial features and BMI, we can actually observe intuitively from some face images about a person's fatness. For example, some face images and the corresponding BMIs are shown in Fig. 1.

Based on the psychology studies and our intuitive observation (as shown in Fig. 1), we believe that it is worth investigating a *computational approach* to body mass index prediction from face images. Furthermore, the computational approach can verify the psychology study results with a large scale of data. For example, in psychology studies, usually a small number of face images are used, thus it is not clear whether the result could hold when more faces are tested. Another thing is that the facial features are usually labeled manually in psychology studies, which constrains the verification to a small scale of data. Finally, the psychology studies [6–8] focus on exploring

the correlations between facial features and BMI, but there is no prediction of BMI from face images. In our computational approach, we will perform BMI prediction and evaluate its accuracy on a large database.

One advantage of predicting BMI from face images is that, the approach is non-invasive. There is no need to measure an individual's height and weight in order to compute his or her BMI. This is a nice property for some practical uses, such as in on-line photos or surveillance videos of faces, where it is impossible to use traditional measures of weight and height for BMI calculation. Recently, there are more and more online dating or friend search sites (e.g., <http://www.onlinedatingsites.net/>), where possibly only face photos are shown for each individual. The automated prediction of BMI from face photos can be useful to judge bodily attractiveness and health [6–8].

1.3. Organization

In the remaining of the paper, we first present our computational approach to BMI prediction in Section 2, including details on feature computation and modeling BMI prediction as a regression problem. Then we introduce the calculation of the Pearson's correlation coefficients between facial features and BMI in Section 3. We present the experiments on a large scale database for BMI prediction and correlation measures in Section 4. And finally, we draw conclusions.

2. A computational approach

We propose a computational approach for BMI prediction from a single face image. The whole framework is shown in Fig. 2. Basically, we formulate the BMI prediction as a machine vision problem. In the following, we describe all related procedures.

2.1. Face detection, alignment, and fitting

For a given face image, the first step is to execute face detection [9], and the detection of two eyes using a technique similar to face detection [9]. The difference is what templates are used for training the AdaBoost classifier. For face detection, the face templates are used, while for eye detection, the templates are replaced with eyes. Then each detected face is normalized based on the detected eye coordinates. The normalization is basically to perform translation, rotation, and scaling of the faces so as to align all face images into



Fig. 1. Some face images with BMI values and categories. Increased facial adiposity can be observed as the BMI increases.

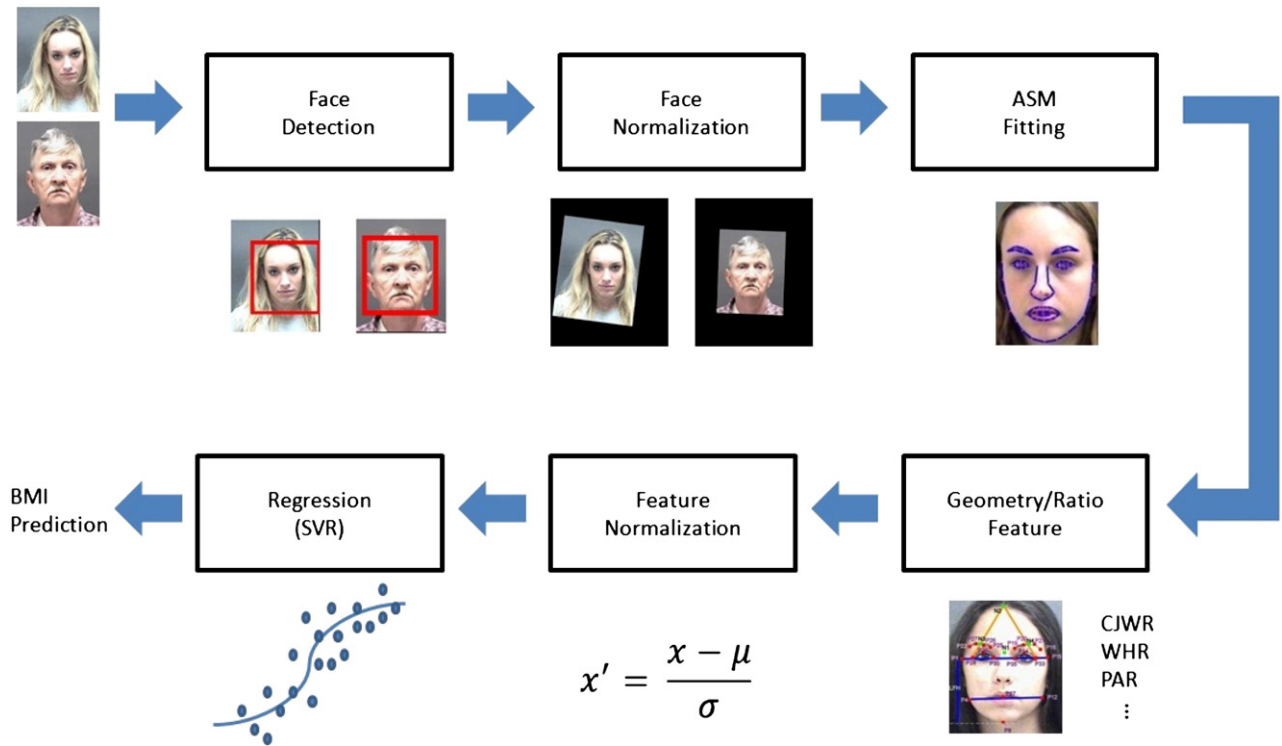


Fig. 2. The proposed framework for body mass index prediction from a face image.

the common eye coordinates. These processing steps are important in developing a robust computational system.

Next, the active shape model (ASM) is used to detect a number of fiducial points in each face image. The ASM method was originally proposed by Cootes et al. [10]. In the ASM model, the principal

component analysis (PCA) was applied on the locations of facial components (e.g., eyes, nose, lips, facial contours, etc.), then ASM was presented as connected point distributions from a variety of manually labeled images with variations such as pose, illumination, and expression changes. Milborrow [11] extends the ASM with a so-called Constrained Local Model to improve its performance. The ASM can perform well in facial feature extraction [12], structure locating in medical images [13], and many other applications. An example of the ASM fitting on a test face image is shown in Fig. 3. Please note that we apply the ASM model to the normalized face images, rather than the original images that may contain various sizes of faces and head rotations. Letting the ASM fitting work on the normalized face images is helpful to make the fiducial point detection more robust and accurate. The detected facial points by the ASM are used as the input for the next step to compute facial geometry or ratio features.

2.2. Facial features

Seven facial features are *automatically* detected or estimated in our approach. They are cheekbone to jaw width (CJWR), width to upper facial height ratio (WHR), perimeter to area ratio (PAR), eye size (ES), lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH), and mean of eyebrow height (MEH). The use of these features is motivated by the psychology studies [7,8]. In our approach, we pursue an automatic detection and estimation of these features, rather than manual labeling as in psychology studies [7,8]. An illustration of these facial features is shown in Fig. 4. In the following, we describe and interpret these features, and then introduce how to compute them automatically.

2.2.1. Seven computed features

There are seven facial features in total. They are visually shown in Fig. 4. The meanings of these features are described below.

- (1). CJWR is the ratio of the cheekbone width to jaw width. The cheekbone width is the distance between P_1 and P_{15} , and jaw



Fig. 3. The result of the ASM fitting on fiducial points in a test face image.

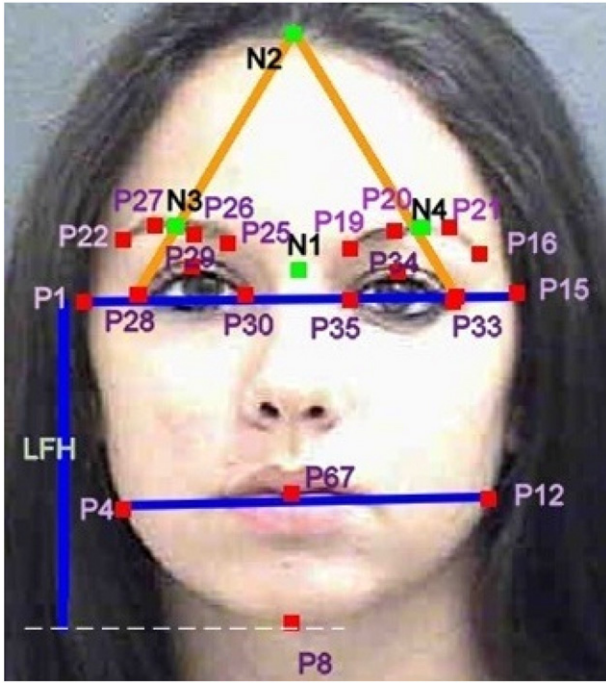


Fig. 4. Illustration of the facial features for BMI prediction.

width is the distance between P_4 and P_{12} . So the CJWR feature is computed by $\frac{|P_1P_{15}|}{|P_4P_{12}|}$;

- (2). WHR is the ratio of the cheekbone width to upper facial height which is the distance between P_{67} and N_1 . The formula of WHR is $\frac{|P_4P_{12}|}{|P_{67}N_1|}$;
- (3). PAR is the ratio of the perimeter to area of polygon running through $P_1P_4P_8P_{12}P_{15}P_1$, so the computation is given by $\text{Perimeter}(P_1P_4P_8P_{12}P_{15}P_1)/\text{Area}(P_1P_4P_8P_{12}P_{15}P_1)$, and the perimeter is computed by the sum of line segment lengths, and the area is the sum of all related triangles, respectively;
- (4). ES is the average size of eyes, which is the average distance between P_{28} and P_{33} minus the distance between P_{30} and P_{35} . The formula is $\frac{1}{2}(|P_{28}P_{33}| - |P_{30}P_{35}|)$;
- (5). LF/FH is the lower face to face height ratio. The lower face is the part of the face below the cheekbone (i.e., line P_1P_{15}). The lower face height is the distance between the cheekbone and the lowest point in the jaw which is shown in Fig. 4 as the distance LFH. The face height is the distance between the highest point N_2 and the lowest point P_8 on the face. The formula is $\frac{|LFH|}{|N_2P_8|}$;
- (6). FW/LFH is the face width to the lower face height ratio. The face width is the cheekbone width. So the FW/LFH feature is computed by $\frac{|P_1P_{15}|}{|LFH|}$;
- (7). MEH is the average distance between eyebrows and the upper edge of eyes, which is the mean of distances between P_{22} to P_{28} , N_3 to P_{29} , P_{30} to P_{25} , P_{19} to P_{35} , N_4 to P_{34} , and P_{16} to P_{33} . Thus the formula will be $\frac{1}{6}(|P_{22}P_{28}| + |N_3P_{29}| + |P_{25}P_{30}| + |P_{19}P_{35}| + |N_4P_{34}| + |P_{16}P_{33}|)$.

2.2.2. Automatic feature computation

In order to compute the features automatically, we use the Active Shape Model (ASM) to detect facial fiducial points first, and then compute the distance and ratio features. The recently developed ASM technique [14] can find 76 points in a face image. We used 20 points out of 76 to produce the facial features that are related to BMI. Those points are marked as P_* in Fig. 4 (* indicates the *th points in the original list returned by the ASM method). Note that even with the big number of 76 points detected by the ASM, as shown in Fig. 3, there are still some other points that are needed to compute the features. Actually, another four points are needed but cannot be detected

by the ASM method, denoted by $N_1 - N_4$ in Fig. 4. We estimate the positions of these four points based on the detected fiducial points by the ASM method.

Specifically, N_1 is estimated by the midpoint of P_{29} and P_{34} , as shown in Fig. 4. N_2 is approximated by the intersection point between lines $P_{28}N_3$ and $P_{33}N_4$. N_3 is estimated by the midpoint of P_{26} and P_{27} , and similarly, N_4 is the midpoint of P_{20} and P_{21} . We use N_2 as an estimate of the highest point on a face. This estimate is based on our observations and a quantitative evaluation that will be presented in the next.

Given the fitted fiducial points by the ASM method, and the estimated points described above, i.e., $N_1 - N_4$, all of the facial geometry and ratio features, introduced in Section 2.2.1, can be computed automatically. These features are supposed to be correlated to BMI, motivated by the psychology studies [7,8]. We believe that it is a significant progress to show that the BMI related facial features can be computed automatically in each face image.

2.3. Normalization

After the computation of facial features using the methods presented above, normalization is executed to get the normalized features by

$$x' = \frac{x - \mu}{\sigma},$$

where μ is the mean value and σ is the standard deviation, computed from the training data along each feature dimension. Since the features may have different ranges of values, we found that the feature normalization is important in order to get a better prediction of the BMI. After feature normalization, we perform a statistical learning to build the relation between facial features and the BMI.

2.4. Statistical learning

Given the computed features in faces, we define the BMI prediction from facial features as a statistical learning problem. We propose to use three regression methods, i.e., the support vector regression (SVR) [15], Gaussian process (GP) [16], and the least squares estimation (LSE) [17], and compare their performance in modeling the relations for BMI prediction. We want to mention that, in psychology studies [6–8], there is no prediction of BMI from facial features. Their studies are mainly the correlation measures between facial features and the BMI.

2.4.1. Support vector regression

Given a set of training data $(x_1, y_1), \dots, (x_l, y_l) \subset X \times R$, where X denotes the space of the input patterns, $X \in R^d$, and y_1, \dots, y_l are the target regression values, e.g., BMI values in our case. extitR denotes one-dimensional real values for the BMI. The SVR [15] is to learn a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. We chose $\varepsilon = 0.002$ in our experiments taking the form

$$f(x) = \langle w, x \rangle + b,$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in X . Flatness in this form means that one seeks a small w . One way to ensure this is to minimize the norm. So this problem can be written as a convex optimization problem [18]

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|w\|^2 \\ & \text{subject to} && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \end{aligned}$$

A quadratic programming is usually used to solve the optimization problem [18]. The solution of the SVR is the w and b .

2.4.2. Least squares regression

Least squares estimation is a traditional method to estimate the relation between two sets of variables. A basic assumption underlying the application of the least squares method is that the error terms in the regression model are independent [17]. Suppose observations are $\{X_i, y_i\}$. The model is

$$y = X\beta + \epsilon,$$

where β is a vector of parameters given by

$$\beta = (X'X)^{-1}X'y,$$

with X' as the transpose of X .

2.4.3. Gaussian process regression

A Gaussian process (GP) [16] is a collection of random variables, any finite number of which has a joint Gaussian distribution. Given a set of training data $(x_1, y_1), \dots, (x_n, y_n) = (X, y)$, the Bayesian analysis of the standard linear regression model with Gaussian noise is given by $f(x) = x^T w$, $y = f(x) + \epsilon$, where x is the input vector, w is a vector of weights of the linear model, f is the function value and y is the observed target value. ϵ is a bias weight or offset. Assuming that the observed values y differ from the function values $f(x)$ by additive noise, and further assuming that this noise follows an independent, identically distributed Gaussian distribution with zero mean and variance σ_n^2 , we get $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. In the Bayesian formalism, we have a zero mean Gaussian prior with covariance matrix Σ_p on the weights $w \sim \mathcal{N}(0, \Sigma_p)$ [16]. To make predictions for a test example x_* , the predictive distribution for $f_* \triangleq f(x_*)$ at x_* is given by averaging the output of all possible linear models with respect to the Gaussian posterior. $p(f_* | x_*, X, y) = \int p(f_* | x_*, w) p(w | X, y) dw = \mathcal{N}\left(\frac{1}{\sigma_n^2} x_*^T A^{-1} X y, x_*^T A^{-1} x_*\right)$, where $A = \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1}$. More details can be referred to [16]. Here we explore the Gaussian process regression for our BMI prediction.

The SVR, GP, and LSE methods are classical regression approaches, and have shown good performance in many applications. However, it is the first time these methods to the problem of BMI prediction are adapted, to the best of our knowledge. In order to understand the new problem of BMI prediction deeply, we apply all three methods and carry on comparisons. We want to investigate how different the performance of BMI prediction is: a direct regression (LSE) or more robust regression, e.g., SVR or GP.

Before performing the BMI prediction experiments, we present the Pearson's correlation measures between facial features and the BMI. We want to understand the correlation on a large database, while the psychology studies [7,8] only use a small number of samples, e.g., hundreds, for correlation measures. Statistically, it is more meaningful to evaluate the correlations on a large database, e.g., more than 10,000 samples.

3. Correlations between facial features and BMI

In addition to BMI prediction, we also examine the correlations between facial features and the BMI. In psychology studies [7,8], the major measure is the Pearson's correlation coefficient, but the correlations are usually measured on a small number of samples, since it is difficult to manually label the facial features on a large database. However, in our computational approach, we can measure the correlations on a large number of samples, which will show more meaningful measures statistically. It can also verify the psychology study results [7,8], but on a much larger number of samples.

Usually the Pearson's correlation coefficient is used to measure the correlation between two sets of variables. We will give a brief

introduction of the Pearson's correlation, and then describe the hypothesis testing, which is used to test the significance of the obtained correlations between facial features and the BMI. Based on the hypothesis testing, we can draw a conclusion on whether the facial features and the BMI are correlated in the case of a large population.

3.1. Pearson's correlation coefficient

Pearson's correlation coefficient is also called Pearson's r . Pearson developed the mathematical formula in 1895. It focuses on the correlation coefficient as a computational index used to measure bivariate association. The correlation coefficient constitutes the principal statistical methodology for observational experiments in many disciplines [19]. It is calculated by

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

The Pearson's correlation coefficient is between -1 and 1 . If $r < 0$, it indicates a negative correlation between X and Y . If $r > 0$, it indicates a positive correlation. If $r = 0$, it indicates there is no correlation between X and Y [20]. If Pearson's r is close to -1 and $+1$, it indicates a very strong correlation.

3.2. Statistical significance

The above Pearson's correlation coefficients are computed from the observed samples. Even if we have a large number of samples for the correlation measure, it is still not equivalent to the correlation measured in the whole population. To extend the correlation measure to the population, we need to do hypothesis testing with a statistical significance measure.

In statistics, p-value describes how certain a hypothesis is true. Here we use the p-value to indicate whether a significant correlation exists between the computed facial features and the BMI. The p-value measures the plausibility of the null hypothesis, which is defined as there is no correlation between facial features and the BMI. The smaller the p-value, the stronger the evidence is against the null hypothesis. If the p-value is sufficiently small, the null hypothesis would be abandoned and an alternate hypothesis is believed [21]. Some threshold values can be set, e.g., 0.001, 0.01 or 0.05. If the p-value is equal or smaller than the thresholds, it indicates a significant correlation between facial features and the BMI.

4. Experiments

Now we perform experimental validations on a large database. We first introduce the database, and then present the BMI prediction results, and the correlation measures between facial features and the BMI.

4.1. Database

The MORPH-II database [22] is used in our research. There are about 55,000 face images in MORPH-II. Because of the imbalance of the age, gender and ethnicity distribution in MORPH-II database, e.g., about 96% face images are the Black and White, but only 4% are Hispanic, Asian, Indian, or others, we only used the White and Black faces in our study. Also considering the gender and age group distributions, we selected part of face images, and divided into two sets: Set1 and Set2. The division into two sets is motivated by the age estimation study in [23]. The details about the distribution can be seen from Table 2. Note that there are 7273 face images in Set1, and 7323 face images in Set2. Among these images, there are 4591 identities in Set1, and 4590 identities in Set2. So there are some individuals

Table 2

The selected data of Set1 and Set2. There are 4 gender and ethnicity groups in both Set1 and Set2: Black female, Black male, White female, and White male. They are also corresponding to six age groups: age10s (<20), age20s (20–29), age30s (30–39), age40s (40–49), age50s (50–59) and age60s (≥ 60).

	Set1	Set2
Black female	908	902
Black male	2642	2692
White female	1006	1022
White male	2717	2707
Age10s	992	993
Age20s	2687	2692
Age30s	1901	1901
Age40s	1243	1251
Age50s	413	455
Age60s	24	20

with more than one face image in either Set1 or Set2. But most of them have different ages and different BMI values, because the MORPH database contains age variations for the individuals. So the influence of those persons on the whole result of BMI prediction should be small in our experiments.

The number of samples used in our experiments is much larger than those used in psychology studies [7,8]. We have the ground truth of weight and height for each individual, thus the true BMI values are known for our evaluation. The range and distribution of the BMI values in Set1 is shown in Fig. 5. The BMI values are mainly in the range of 15 to 35, which contains all BMI categories: underweight, normal, overweight and obese, as shown in Table 1. From Fig. 1, we can also get a visual perception of different BMI values in the faces.

4.2. On facial points

Our goal is to perform BMI prediction from face images automatically. To achieve the goal, one major step is to detect facial points automatically for feature extraction. In our current approach, most facial points can be detected automatically by the ASM method, however, there are some points that cannot be obtained by the ASM. As stated in Section 2.2.1, some facial points are predicted or estimated by the other ASM detected points, rather than by manual labeling. It is not practical or convenient to ask the users to manually label some facial points, especially in test face photos. To have a brief evaluation of our

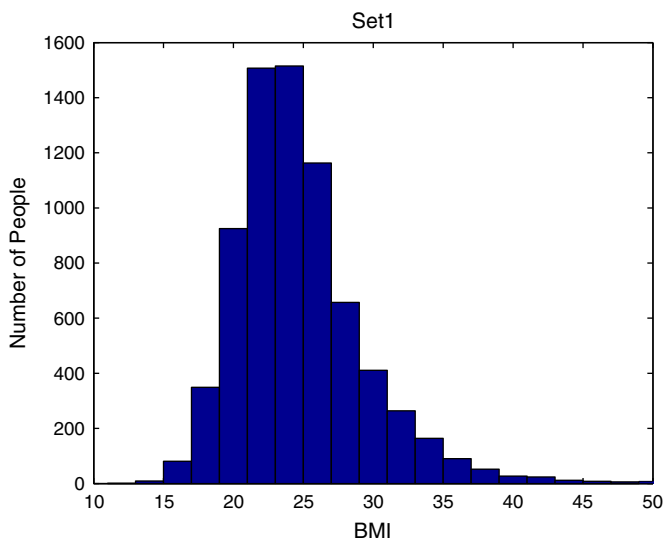


Fig. 5. The distribution of BMI values on a large database (in Set1). The BMI values span a wide range with most of the values between 15 and 35. The distribution and BMI range are similar in Set2 and are not shown here.

estimation method for the facial points, especially the highest point of a face, N_2 , we randomly selected 100 images from our experimental database – MORPH-II [22], and manually labeled all the needed points to produce N_2 and also the real highest point of each face. Then we found the average deviation of our estimate w.r.t. the manually labeled positions of the highest points is about 20.19 pixels. While the average face size of the 100 images is about 182×230 pixels, this deviation is about 8.8% of the face height. We consider this deviation is small for our automated feature extraction. Moreover, the BMI prediction errors are not large (see results in this section), which indirectly show the feasibility of estimating the point location for feature extraction.

Next, we present the experimental results, including both correlation measures and the BMI prediction performance analysis.

4.3. Correlations between facial features and BMI

The correlations between facial features and the BMI are measured and shown in Table 3. One can see that all of the correlations are significant based on the extremely small p-values in both data sets. This indicates that the extracted facial features are correlated to the BMI statistically.

Then we investigate the correlations between facial features and the BMI along with the age, gender, and ethnicity. We use the data in Set1 for this investigation. The results are shown in Tables 4 and 5. In these tables, p-value = 0.0000 indicates a very small p-value, which is less than 0.00005.

In the following discussion, we consider correlation with p-value ≤ 0.05 as a significant correlation. Smaller CJWR and larger WHR indicate a wider and squarer face [8]. Smaller PAR indicates more round low face which also means a wider and squarer face. Smaller ES indicates smaller eyes. Smaller LF/FH indicates smaller rate between lower face to whole face length. Larger FW/LFH also indicates a wider face. Since usually, the height of people's face is larger than the width, we consider a wider face as a squarer face.

Table 4 shows the correlations between the facial features and the BMI in six age groups. We can observe the tendency between the facial features and the BMI in different groups. In the age10 group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH) and smaller eyes (ES) has a larger BMI. In the age20 group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eyes (ES), smaller rate between lower face to whole face length (LF/FH), larger eyebrow height (MEH) has larger BMI. In the age30 group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller rate between lower face to whole face length (LF/FH), and smaller eyes (ES) has larger BMI. In the age40 group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller rate between lower face to whole face length (LF/FH), smaller eyes (ES), and larger eyebrow height (MEH) has bigger BMI. In the age50 group, a person with a squarer face (CJWR, WHR, and FW/LFH) has larger BMI. There is no feature indicating significant correlations with BMI in the age60 group. This may be caused by the insufficient number of samples

Table 3

Pearson's correlation coefficient r between facial features and BMI on Set1 and Set2, and the corresponding p-values.

	Set1		Set2	
	r	p-value	r	p-value
CJWR	−0.20	0.0000	−0.20	0.0000
WHR	0.25	0.0000	0.28	0.0000
PAR	−0.07	0.0000	−0.04	0.0000
ES	−0.10	0.0000	−0.11	0.0000
LF/FH	−0.07	0.0000	−0.09	0.0000
FW/LFH	0.15	0.0000	0.19	0.0000
MEH	0.04	0.0002	0.05	0.0001

Table 4
Pearson's r between facial features and the BMI in age groups (in Set1).

	Age10s		Age20s		Age30s		Age40s		Age50s		Age60s	
	r	p-value	r	p-value	r	p-value	R	p-value	r	p-value	r	p-value
CJWR	−0.16	0.0000	−0.21	0.0000	−0.21	0.0000	−0.21	0.0000	−0.11	0.0213	−0.22	0.3046
WHR	0.27	0.0000	0.24	0.0000	0.27	0.0000	0.26	0.0000	0.23	0.0000	0.16	0.4622
PAR	−0.10	0.0013	−0.10	0.0000	−0.08	0.0002	−0.07	0.0163	0.07	0.1631	−0.09	0.6870
ES	−0.09	0.0050	−0.07	0.0001	−0.09	0.0000	−0.10	0.0003	−0.08	0.1249	0.06	0.9412
LF/FH	−0.05	0.0956	−0.10	0.0000	−0.08	0.0007	−0.10	0.0005	−0.04	0.4727	−0.20	0.3456
FW/LFH	0.19	0.0000	0.15	0.0000	0.16	0.0000	0.11	0.0000	0.24	0.0000	0.06	0.7845
MEH	0.05	0.0981	0.09	0.0000	0.04	0.1012	0.07	0.0139	−0.10	0.0598	0.15	0.4853

used in this age group. So it is important to use a large number of samples for analysis.

Table 5 shows the correlations between facial features and BMI in four gender–ethnicity groups. In the White female group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH) has larger BMI. In the White male group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eye size (ES) has larger BMI. In the Black female group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), larger eyebrow height (MEH) has larger BMI. In the Black male group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eyes (ES) has larger BMI.

A comparison between the correlation measures in our work and the psychology study results is given in Table 6. Our results have similar correlations with those in [8,7], except the ES. Maybe that is because in [8], the participants are the Korean people, different from the ethnicity groups in our database.

Overall, the correlation measures on our database show that most of the facial features are related to the BMI, although some correlations exist with different results for different age, gender, or ethnicity groups. Based on this, we use the facial features to predict the BMI, which will be presented next.

4.4. BMI prediction from faces

We propose to measure the performance of BMI prediction by the mean absolute error (MAE). The MAE is defined as the average of the absolute errors between the estimated BMIs and the ground truth BMIs. $MAE = \frac{1}{N} \sum_{k=1}^N |b_k - \hat{b}_k|$, where b_k is the ground truth BMI for the test image k , \hat{b}_k is the estimated BMI for image k , and N is the total number of test images. This measure is motivated by that used in age estimation, e.g., [24].

Table 7 shows the MAEs of BMI prediction using the regression methods, SVR (with the RBF kernel), GP, and LSE. In Table 7, the MAEs are shown in different age groups. We used the Set1 for training, while Set2 for testing, or vice versa. The MAEs of BMI is among a range of 3.00 ± 1.00 . This indicates that the error of the BMI prediction is relatively small, comparing to the wide range of BMI from 15 to 35, visually shown in Fig. 5. As shown in Table 7, the MAEs are slightly different between different groups. For example, the age10 group has a smaller MAE than other age groups.

Table 5
Pearson's correlation coefficient r between facial features and the BMI in gender–ethnicity groups (in Set1).

	White female		White male		Black female		Black male	
	R	p-value	r	p-value	r	p-value	R	p-value
CJWR	−0.25	0.0000	−0.19	0.0000	−0.26	0.0000	−0.14	0.0000
WHR	0.21	0.0000	0.25	0.0000	0.30	0.0000	0.20	0.0000
PAR	−0.19	0.0000	−0.07	0.0006	−0.22	0.0000	−0.06	0.0016
ES	0.01	0.7500	−0.10	0.0000	0.04	0.1900	−0.05	0.0117
LF/FH	−0.03	0.4200	−0.02	0.2000	−0.06	0.0600	−0.03	0.1400
FW/LFH	0.13	0.0001	0.22	0.0000	0.14	0.0000	0.15	0.0000
MEH	−0.00	1.0000	−0.03	0.1200	0.07	0.0400	−0.01	0.7600

Fig. 6 shows the MAEs in the whole Set1 or Set2. The left shows the MAEs when training in Set2 while testing in Set1. The right shows the MAEs when training in Set1 while testing in Set2. From this figure, we can observe that overall the SVR method performs better than both the GP and LSE methods.

Fig. 7 shows the MAEs of SVR, GP and LSE in different BMI categories: underweight, normal, overweight, and obese. We can see that all the three methods perform well in the normal and overweight BMI categories. In the normal BMI categories, the MAE of SVR is only 1.91. As shown in Table 8, the normal BMI category contains about 55% of the faces. The overweight category contains about 30%. These two categories possess about 85% of the whole set. As shown in Fig. 7, in the normal BMI category, the MAE of SVR is the smallest. In the overweight category, the MAE of GP is the smallest. As a result, we can get a higher accuracy in predicting the BMI according to facial features in these two categories. The lower accuracies in another two categories, i.e., underweight and obese, are probably because of the smaller number of samples for both training and testing, as shown in Table 8. The result suggests that a large number of training examples are needed in order to have a better BMI prediction performance.

Even though the SVR, GP and LSE have similar MAE results, the SVR is better for the underweight and normal BMI categories, while the GP performs better for overweight and obese BMI categories. Overall the SVR has the smallest errors as shown in Fig. 6.

4.5. Other features

In our study, the facial features that we used are motivated by the observations in psychology and human perception studies [6–8]. One may argue that probably some other features could be used too. One possible choice is the ASM shape feature [10], which is popular for facial shape analysis and face recognition. To get an understanding on this, we use the ASM feature for facial feature characterization for BMI prediction, using the SVR for regression. Experimentally, we found that the MAE of the ASM shape feature is 3.34, based on the average of Set1 to Set2 and Set2 to Set1. This MAE is slightly higher than the 3.14 for the seven features that we used and motivated by

Table 6
Correlations between facial features and BMI: A comparison. The age20s group is selected to show, since this age group appeared in [7,8]. An empty cell means no result available, and “#” indicates the corresponding p-value > 0.05.

	Coetzz et al. [7]		Pham et al. [8]		Our work	
	Male ($n = 187$)	Female ($n = 194$)	Male ($n = 230$)	Female ($n = 229$)	Male ($n = 1973$)	Female ($n = 714$)
PAR	−0.22	−0.12#	−0.30	−0.23	−0.16	−0.21
WHR	0.17	0.36	0.30	0.28	0.29	0.27
CJWR	−0.20	−0.29	−0.40	−0.29	−0.24	−0.26
ES			0.29	0.26	−0.07#	−0.05#
LF/FH			0.12#	0.10#	−0.02#	−0.02#
FW/LF			0.12#	0.05#	0.20	0.15
MEH			0.05#	0.16	−0.04#	0.01#

Table 7

The BMI prediction errors measured by the mean absolute error (MAE) on a large database. The results are presented based on age groups and in total average. "Set1 Set2" means the data in Set1 is used for training, while Set2 is used for testing. Three different regression methods, SVR, GP, and LSE, are used to learn the BMI prediction function.

	Set1 Set2			Set2 Set1		
	SVR	LSE	GP	SVR	LSE	GP
Age10s	2.79	3.01	3.02	2.64	2.90	2.89
Age20s	3.03	3.11	3.15	3.08	3.19	3.18
Age30s	3.47	3.44	3.47	3.42	3.42	3.41
Age40s	3.13	3.17	3.21	3.20	3.21	3.18
Age50s	3.24	3.41	3.26	3.21	3.22	3.19
Age60s	2.52	2.23	2.23	2.66	2.75	2.59
Avg.	3.14	3.21	3.23	3.14	3.22	3.20

the human perception studies [6–8]. This comparison shows that the features that we used are capable of characterizing the facial property better than the ASM features for BMI prediction. Our interpretation is that the ASM shape parameters characterize the generic shape features and pertain to the individual's shape, and thus may not deliver the "salient features" pertaining to the BMI prediction.

Based on our exploration, it may inspire some future investigations to develop more discriminative features that can characterize and predict the BMI with a better performance.

4.6. BMI prediction in separate gender and ethnicity groups

Finally, we study the BMI prediction performance when it is executed in separate gender and ethnicity groups. In this study, the training and test faces come from the same gender and ethnicity groups. The same group in Set1 is used for training, and the data in Set2 is used for testing, and vice versa. Since the SVR method performs the best overall, we use the SVR for this experiment. The results are shown in Table 9. For comparison, we also show the BMI prediction results using mixed training data, i.e., without separation of the four groups in training, in Table 10.

From Tables 9 and 10, we can observe that the MAE differences are small between the separated training and mixed training, although the MAE can be reduced in some cases, when the separated training is executed. So there is no need to perform a separate training for BMI prediction based on our current experiments. Note that in order to do the BMI computation in separated groups, the gender

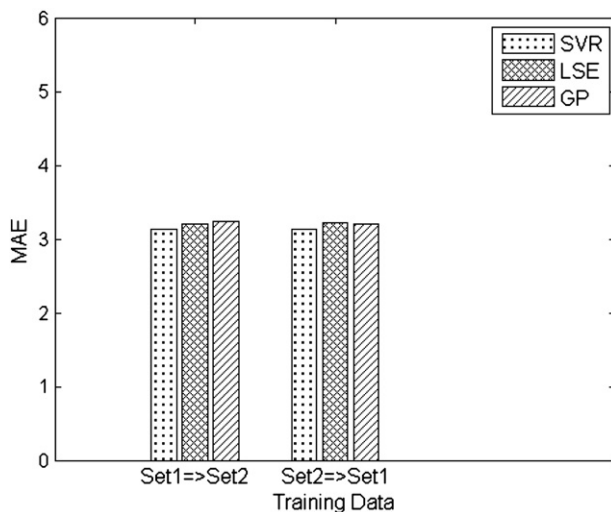


Fig. 6. Visualization of the MAEs of SVR, GP and LSE in total average. Left: Set1 Set2; Right: Set2 Set1. Overall, the SVR gives the smallest errors among the three methods.

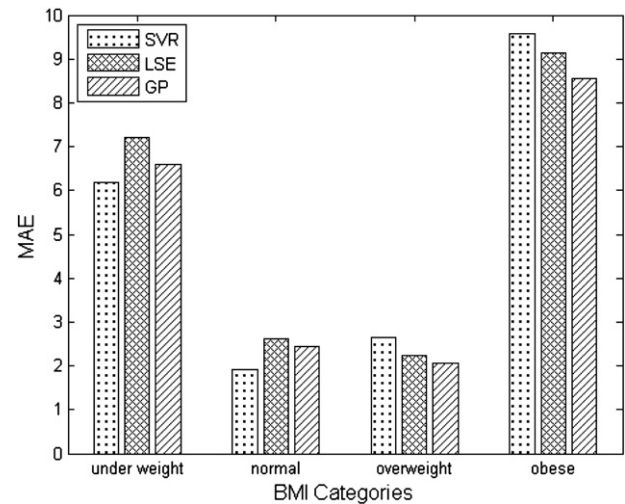


Fig. 7. Visualization of the MAEs of SVR, GP, and LSE in different BMI categories. The SVR has the lowest errors for underweight and normal, while the GP has the smallest errors for overweight and obese categories.

and ethnicity information has to be recognized first, which is beyond the scope of this paper.

From Tables 9 and 10, we can also observe that the BMI prediction errors are different for different groups of people. For example, the MAEs for the male are small, while the errors for the female are large. This result suggests that our computational BMI prediction is more accurate for males than for females.

5. Concluding remarks

We have developed an automated, computational system for body mass index prediction from face images. It is motivated by the recent

Table 8

This table shows the number of face images in four BMI categories. The normal BMI category contains the most number of faces, which is about 55% of the whole set. Overweight category is the second largest, which contains about 30% of faces.

	Set1	Set2
Underweight	308	303
Normal	4007	4042
Overweight	2096	2093
Obese	838	862

Table 9

The BMI prediction errors (measured by the MAE) based on separated training in each of the four groups, using the SVR method.

	Set1 Set2	Set2 Set1
Black female	4.18	3.91
Black male	3.07	3.16
White female	3.44	3.22
White male	2.69	2.78
Average	3.13	3.12

Table 10

The BMI prediction errors (measured by the MAE) based on the SVR method using mixed training (without gender and ethnicity separation). The results are presented for the four groups separately.

	Set1 Set2	Set2 Set1
Black female	4.29	3.94
Black male	3.05	3.13
White female	3.64	3.44
White male	2.65	2.76
Average	3.14	3.14

studies in psychology and human perception. It also validates the psychology study results on a large database with more than 14,500 face images. The correlation coefficients and p-value measures in the large database demonstrate the correlations between the computed facial features and the BMI in a statistically meaningful manner. We have shown that a computational approach can be developed for BMI prediction using machine vision and statistical learning techniques. Our developed system is probably the first computational approach to predict BMI from face images automatically.

In future work, we will explore more facial features for BMI prediction, and study age, gender, or ethnicity group-specific features to better characterize the facial appearance for the purpose of BMI prediction.

Acknowledgment

The authors would like to thank K. Ricanek for providing the MORPH database for this study. The work is partially supported by an NSF CITEr grant and an NIJ grant 2010-DD-BX-0161. The authors are grateful to the anonymous reviewers for their detailed comments and suggestions to improve the paper.

References

- [1] L. Zebrowitz, Finally, faces find favor, *Soc. Cogn.* 24 (5) (2006) 657–701.
- [2] D. Gallagher, M. Visser, D. Sepulveda, R. Pierson, T. Harris, S. Heymsfield, How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? *Am. J. Epidemiol.* 143 (3) (1996) 228–239.
- [3] A.G. Renehan, M. Tyson, M. Egger, R.F. Heller, M. Zwahlen, Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies, *Lancet* 2008 (371) (2008) 536–546.
- [4] A. Keys, F. Fidanza, M.J. Karvonen, N. Kimura, H.L. Taylor, Indices of relative weight and obesity, *J. Chron. Dis.* 25 (6–7) (1972) 329–343.
- [5] N. Sebire, M. Jolly, J. Harris, J. Wadsworth, M. Joffe, R. Beard, L. Regan, S. Robinson, Maternal obesity and pregnancy outcome: a study of 287,213 pregnancies in London, *Int. J. Obes.* 25 (8) (2001) 1175–1182.
- [6] V. Coetzee, D.I. Perrett, L.D. Stephen, Facial adiposity: a cue to health? *Perception* 38 (2009) 1700–1711.
- [7] V. Coetzee, D.I. Perrett, L.D. Stephen, Deciphering faces: quantifiable visual cues to weight, *Perception* 39 (2010) 51–61.
- [8] P. Duong Duc, D. Jun-Hyeong, K. Boncho, L. Hae Jung, K. Honggie, K. Jong Yeol, Body mass index and facial cues in sasang typology for young and elderly persons, *Evidence-Based Complementary and Alternative Medicine*, 2011.
- [9] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple, *Proc. IEEE CVPR*, 2001.
- [10] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, et al., Active shape models – their training and application, *Comput. Vis. Image Underst.* 61 (1) (1995) 38–59.
- [11] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, *Europe Conf. Computer Vision*, 2008, pp. 504–513.
- [12] M.H. Mahoor, M. Abdel-Mottaleb, Facial features extraction in color images using enhanced active shape model, *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 5–11.
- [13] T.F. Cootes, A. Hill, C.J. Taylor, J. Haslam, Use of active shape models for locating structures in medical images, *Image Vis. Comput.* 12 (6) (1994) 355–365.
- [14] S. Milborrow, Locating facial features with active shape models, Ph.D. thesis, Faculty of Engineering, University of Cape Town (2007).
- [15] V.N. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.
- [16] C. Rasmussen, C. Williams, *Gaussian Processes for Machine Learning*, vol. 1, MIT Press, Cambridge, MA, 2006.
- [17] J. Durbin, G.S. Watson, Testing for serial correlation in least square: I, *Biometrika* 37 (3/4) (1950) 409–428.
- [18] A. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (2004) 199–222.
- [19] J.L. Rodgers, W.A. Nicewander, Thirteen ways to look at the correlation coefficient, *Am. Stat.* 42 (1) (1988) 59–66.
- [20] R. Wilcoxon, *Fundamentals of Modern Statistical Methods Substantially Improving Power and Accuracy*, Springer, 2010.
- [21] W. Navidi, *Statistics for Engineers and Scientists*, 2nd ed. McGraw-Hill, 2008.
- [22] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, *IEEE Conf. on AFGR*, 2006, pp. 341–345.
- [23] G.-D. Guo, G. Mu, Human age estimation: what is the influence across race and gender? *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2010.
- [24] G.-D. Guo, G. Mu, Y. Fu, T.S. Huang, Human age estimation using bio-inspired features, *IEEE Conference on Computer Vision and, Pattern Recognition*, 2009, pp. 112–119.