Pattern Anal Applic (2010) 13:289–300
DOI 10.1007/s10044-009-0155-0

THEORETICAL ADVANCES

A template-based approach to automatic face enhancement

Stefano Melacci · Lorenzo Sarti · Marco Maggini · Marco Gori

Received: 30 July 2008 / Accepted: 4 February 2009 / Published online: 8 April 2009

Abstract This paper presents Visual ENhancement of USers (VENUS), a system able to automatically enhance male and female frontal facial images exploiting a database of celebrities as reference patterns for attractiveness. Each face is represented by a set of landmark points that can be manually selected or automatically localized using active shape models. The faces can be compared remapping the landmarks by means of Catmull–Rom splines, a class of interpolating splines particularly useful to extract shape-based representations. Given the input image, its landmarks are compared against the known beauty templates and moved towards the *K*-nearest ones by 2D image warping. The VENUS performances have been evaluated by 20 volunteers on a set of images collected during the Festival of Creativity, held in Florence, Italy, on October 2007. The experiments show that the 73.9% of the beautified faces are more attractive than the original pictures.

Keywords Face analysis · Beautification · Attractiveness · Facial feature localization · Image warping

S. Melacci · L. Sarti (✉) · M. Maggini · M. Gori
Department of Information Engineering, University of Siena, Via Roma, 56, 53100 Siena, Italy
e-mail: sarti@dii.unisi.it

S. Melacci
e-mail: mela@dii.unisi.it

M. Maggini
e-mail: maggini@dii.unisi.it

M. Gori
e-mail: marco@dii.unisi.it

1 Introduction

The perception of human facial attractiveness has fascinated the scientists for centuries and has been investigated from different points of view. The common notion suggests that "beauty stays in the eye of the beholder", and standards of beauty are culturally specific. However, the research activity in this field has made progresses over the last two decades, facing the problem from the psychological, medical, and statistical points of view. The analysis of cognitive and psychological studies shows surprising results that partially confute common assertions about attractiveness. In particular, many experiments were carried out using volunteers, chosen among different races. This cross-cultural analysis suggests that faces of different ethnic groups possess similar structural features perceived as attractive regardless of the racial and cultural background of the perceiver. Many studies were performed in order to understand the role of the average faces and symmetry in the perception of beauty. In the psychological field, the researchers are unanimous in considering that an average face, computed starting from a set of random faces, is more attractive than the original ones [15]. At the same time, Cunningham et al. [1] proved that, even if averaged faces are attractive, a very beautiful face is not close to this average. Also the role of symmetry has been explained, showing that an asymmetrical face is perceived as more attractive than a symmetrical one. In fact, Swaddle et al. [24] performed some experiments using an automatic system able to transform asymmetrical faces into symmetrical ones by means of mirroring and morphing. The original and transformed images were evaluated by some supervisors that generally preferred the asymmetrical faces. Other researchers have focused their efforts to explain how the skin appearance affects the perception of attractiveness and to understand whether some face parts are more important than others to assess the beauty of a face. Fink et al. [10] showed that the attractiveness ratings made by volunteers are strictly related to the skin homogeneity. Finally, Michiels [22] proved that the chin, upper lip, and nose have a great effect on the overall judgment than other face parts.

From a statistical point of view, some scientists followed an anthropometric approach, in order to define a set of facial measurements that predicts the attractiveness ratings [9], but the reported results are often contradictory, suggesting that the answer to the question of what is perceived as beauty cannot be explained using only anthropometric measurements. However, in a recent work, Jefferson [14] proved that beautiful faces have an ideal facial proportion (*divine proportion*) and such proportion is strictly related to the well-known *golden ratio* (1.61803...). Jefferson proved his theory by examples, arguing that there are possibly billions of examples of golden ratio occurrences within the human body.

In the last few years, some automatic systems have been developed with the aim of learning to rate the beauty of a face or to beautify a given facial image. Piccardi et al. [13] presented a system able to analyze frontal facial images and predict their beauty by means of supervised learning. This technique exploits the assumption that an attractive face is characterized by a set of measurements that are related to the golden ratio. Hence, the system represents each image using a set of features that measure the similarity of the faces with respect to the divine proportion. The system is trained to predict the beauty classes defined on a learning set by human supervisors. Three kind of classifiers were evaluated: decision-trees, multi-layered perceptrons and kernel density estimators. The results presented by Piccardi show, at the same time, that the evaluations associated by the supervisors are homogeneous and the attractiveness is somehow related to the golden ratio. Another system able to predict the beauty of a face is described by Eisenthal et al. [8]. Also this system exploits the ratings of some supervisors, assuming that individuals rate facial attractiveness according to similar universal standards. Each face is represented using a set of landmark points and a set of features that describe the relative positions of such points. The supervisors associate both a class membership (attractive/non-attractive) and a rating that ranges in [1,45] to each image. Then, SVMs are trained to solve both the classification and regression tasks. The results are encouraging and suggest that a universal concept of beauty exists.

Finally, four systems able to automatically beautify an input facial image have been recently presented. The first one, proposed by Arakawa et al. [2], is based on the assumption that homogeneous skin is generally more attractive. Therefore, this method exploits a set of image filters able to reduce face imperfections, like wrinkles and moles. The authors presented some processed images that allow us to qualitatively evaluate the results; however, the initial assumption about the homogeneity of the skin can be considered too restrictive, since the perception of attractiveness is affected also by the shape of the face parts and by their relative positions. The second one, presented by Liu et al. [17], automatically beautifies face portraits, replacing the original background with a virtual one and by altering the skin color of the subjects by means of color temperature estimation. The third approach, proposed by Liu et al. [18], is designed to improve the image quality for video conferencing. Interestingly, in this work an "appealing color model" is estimated using a set of face images that depict celebrities. These pictures are assumed to be taken by professional photographers who are able to correctly set up both the brightness and the color tone of the images. Then, during a video conference, for each frame, the color of the face region is changed, in order to move it toward the estimated appealing color model. Finally, the fourth system, presented by Leyvand et al. [16], increases the attractiveness rating of female facial images, and is based on the beauty prediction method proposed by Eisenthal et al. [8]. The key component of this system is a support vector regressor, trained to associate attractiveness ratings to facial images. An input image is represented using a set of landmarks from which a point in a high dimensional "face space" is computed using a set of distances between them. Given an input image, its position in the face space is moved by using the potential field defined by the regressor to increase the associated rating. Then, the corresponding 2D warp is applied to enhance the input image.

In this paper, a template-based approach to automatically enhance both male and female frontal facial images is presented. The study has given rise to the VENUS system, that exploits a database of faces that are commonly considered particularly beautiful (e.g. actors, TV stars, etc.) in order to define its concept of beauty. The faces are represented using a set of 49 landmark points. These points have been manually placed on the learning set, while they are automatically found on input images using active shape models (ASMs) [6]. In order to make faces comparable, the landmark positions are remapped using Catmull–Rom splines (CRSs) [5]. Given an input image, the system compares the automatically extracted features against the known beautiful faces and moves the landmark points of the input image towards the landmarks of the *K*-nearest templates, by a 2D image warping. The system has been recently tested during the Festival of Creativity[1], a public

[1] http://www.festivaldellacreativita.it

event held in Florence, Italy. The collected results have been evaluated by 20 volunteers who have judged, for each pair of images (the original and the beautified face), which one is preferable. The results show that in most cases (73.9%) the beautified faces are considered more attractive than the original ones. Potential applications of the system are in the entertainment and cosmetic industry, virtual media and plastic surgery.

This paper is organized as follows. In the next section, both CRSs and the face representation model will be presented. In Sect. 3 the system architecture will be described, while Sect. 4 collects the experimental results. Finally, in Sect. 5, some conclusions will be drawn.

2 Face representation

The face parts are represented using a set of control points (landmarks). When exploiting a face feature localization technique or manually selecting the landmarks, no guarantee exists that the description of the contours is represented by equally spaced landmarks along the boundaries of the face parts. Since the landmarks do not correspond among different faces, and their number is not predefined, the face parts that they describe are really difficult to compare (see Fig. 1a).

In the general shape matching framework, defining correspondences between shapes generally leads to high computational costs [3]. To overcome this problem, a standard face representation model, defined by a set of landmarks that can be extracted from an incomplete and not continuous description of the boundaries of the face parts, is proposed [19]. Such landmarks can be computed interpolating the initial control points by means of CRSs. CRSs belong to the class of cubic interpolating splines with a simple piecewise construction. Each segment of a CRS interpolates the curve between two control points $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i+1}$ but it is defined by four control points $\boldsymbol{p}_{i-1}$, $\boldsymbol{p}_i$, $\boldsymbol{p}_{i+1}$, $\boldsymbol{p}_{i+2}$. In matrix form, a CRS is described as

$$\boldsymbol{q}(t) = \frac{1}{2}\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{p}_{i-1} \\ \boldsymbol{p}_i \\ \boldsymbol{p}_{i+1} \\ \boldsymbol{p}_{i+2} \end{bmatrix}, \tag{1}$$

where $t \in [0,\ 1]$ describes the curve from $\boldsymbol{p}_i$ ($t = 0$) to $\boldsymbol{p}_{i+1}$ ($t = 1$). Moreover, the tangents at points $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i+1}$ are computed as

$$\boldsymbol{d}_i = \frac{1}{2}(\boldsymbol{p}_{i+1} - \boldsymbol{p}_{i-1})$$

$$\boldsymbol{d}_{i+1} = \frac{1}{2}(\boldsymbol{p}_{i+2} - \boldsymbol{p}_i)$$

Given a set of n control points, with $n \geq 4$, a CRS has C^1 continuity, offers local control, meaning that a change in the position of a control point does not require to recompute the whole curve, and interpolates all the points. The last property is always true in the case of closed curves, otherwise the first and the last control points are excluded from the open curve. Anyway, those two points can be added to the curve by defining the starting and ending tangents. CRSs can exploit an unbounded number of control points to describe either closed or open curves. A single CRS segment is defined by the parameter t, that ranges in [0,1] (see Eq. 1). Unfortunately, equally spaced samples in the parameter space are not equally spaced along the curve. To overcome this limitation, CRSs are redefined using *arc-length parameterization* [25]. The normalized arc-length parameter $s \in [0,\ 1]$ is defined as a function of t, $s = L(t)$, where L is a bijective and monotonically increasing function. Hence, a CRS segment $\boldsymbol{q}(t) = [x(t),\ \ y(t)]$ with $t \in [0,\ \ 1]$ can be expressed with respect to s as $\boldsymbol{r}(s) = [x(L^{-1}(s)), y(L^{-1}(s))]$, where $s \in [0,\ 1]$. The arc length from 0 to t can be computed as $L(t) = \int_0^t \sqrt{(x'(v))^2 + (y'(v))^2}\,\mathrm{d}v$.

Since t ranges between 0 and 1, the length of the whole segment is $L(1)$. In order to find the spline point that corresponds to a given arc-length $s^{\star}$, the function $t = L^{-1}(s^{\star})$ must be numerically approximated.

Considering a CRS that interpolates a set of n control points, the whole curve, that is piecewise-cubic, can be

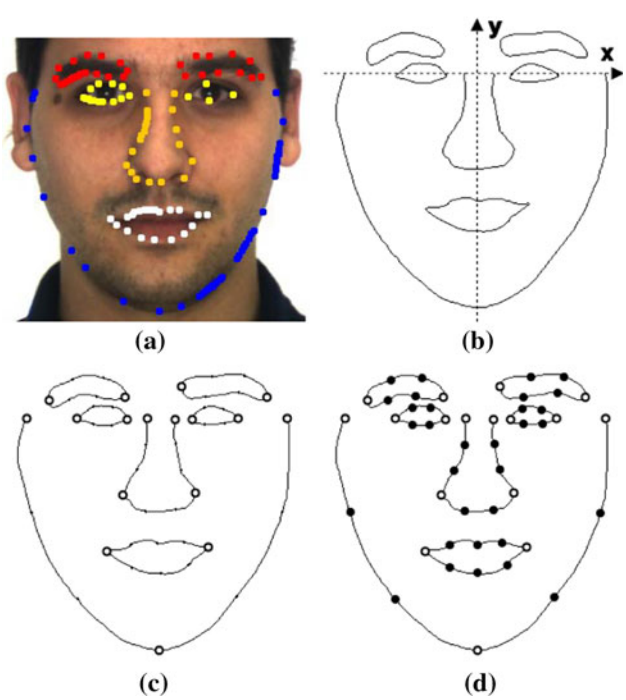

Fig. 1 Face representation extraction: **a** Localized face parts, **b** CRS interpolation, **c** Principal points, **d** Principal and secondary points

described using a global arc-length parameter $gs \in [0, 1]$. If gs_i is its value at the control point i, it can be computed recursively as

$$gs_0 = 0$$
$$gs_i = gs_{i-1} + \frac{L_i(1)}{\sum_{j=1}^{T} L_j(1)}, \quad \text{with } i = 1, \ldots, (n-1),$$

where $L_i(1)$ is the arc-length of the segment between $i-1$ and i, while $T = n$ or $T = n - 1$ for closed or open curves, respectively.

Given a frontal view facial image, its representation can be obtained by:

- an automatic or manual extraction of sets of landmarks that belong to the boundaries of face parts. The coordinates of the landmarks are represented using a reference frame horizontally aligned with the eyes and centered between them;
- the interpolation of each set, using CRSs;
- the localization of a set of *principal points* (PPs) along each CRS, computed using a priori knowledge about the face structure;
- the computation of a fixed number of *secondary points* (SPs).

Given the coordinates of a set of landmarks belonging to the boundary of a face part, they must be ordered CCW or CW and interpolated using a CRS, to obtain a description of the whole contour by means of a parametric curve (see Fig. 1b).

The proposed representation considers the following face parts: face, nose, mouth, eyes and eyebrows (see Fig. 1). The feature points that define this model are extracted from the CRSs that describe each part. PPs are located using a priori knowledge on the structure of the face (see Fig. 1c) and represent salient face features. One-to-one correspondences between PPs belonging to different faces can be defined, thanks to the common structure of the faces themselves. Moreover, PPs can be easily computed since they correspond to spline maxima or minima along a coordinate.

Secondary points are the second type of points constituting the face representation model (see Fig. 1d) and are equally spaced between two consecutive PPs. Equally spaced points along the contours are used in shape modeling and statistical shape analysis [7], and constitute a quite accurate method for defining a set of corresponding points between shapes. Since SPs are defined between two consecutive PPs, the presence of noise in a spline portion could affect the positioning of equally spaced points, but it has no effects on the other parts of the curve. This property makes the face representation model more robust against errors in the segmentation process than shape description approaches based on the definition of just the starting point for parametric curves. Given an arc-length parameterization of the whole CRS, an arbitrary number m of SPs between PP_1 and PP_2, with associated global spline parameter gs_1 and gs_2 ($gs_1 < gs_2$), can be computed by a uniform sampling of the spline segment between PP_1 and PP_2.

The proposed representation technique exploits the reference frame to handle in-plane head rotations. Instead, out-of-plane head rotations should be avoided, since they produces an artifact asymmetrical representation of the facial features.

The proposed representation can be fast and easily computed, and can be used in every task that needs to compare faces. Furthermore, it can lead to a compact parametric approximation of the contours of face parts, by discarding the initial landmarks and considering the CRSs that interpolate the sets of PPs and SPs. In this case, the number of SPs must be chosen in order to find a trade-off between the approximation accuracy and the representation compactness.

The choice of the number of SPs represents a parameter for the beautification system. Currently, each face is represented using 17 PPs and 32 SPs, but there is no constraint on the number of SPs that can be extracted from each spline segment.

3 The VENUS system

The VENUS system exploits a set of supervised examples to define the concept of face beauty. As shown in Fig. 2, the beautification process starts with the extraction of an internal representation of faces by means of CRSs, as described in Sect. 2.

This representation allows the system to compare the input image with respect to the beautiful face templates. Then, the VENUS engine computes the changes that have to be applied to the original image. This phase can be tuned by setting some parameters that control the degree of enhancement. At last, an image warping is performed to obtain the final result. Whereas the landmarks for the training set are manually placed to reduce the problems due to the potential inaccuracies of the automatic extraction, the landmarks of an input image are localized by means of the active shape model algorithm.

3.1 Beautiful face database

The most challenging task of the beautification process is the definition of the beauty concept used by the system to improve the face attractiveness.

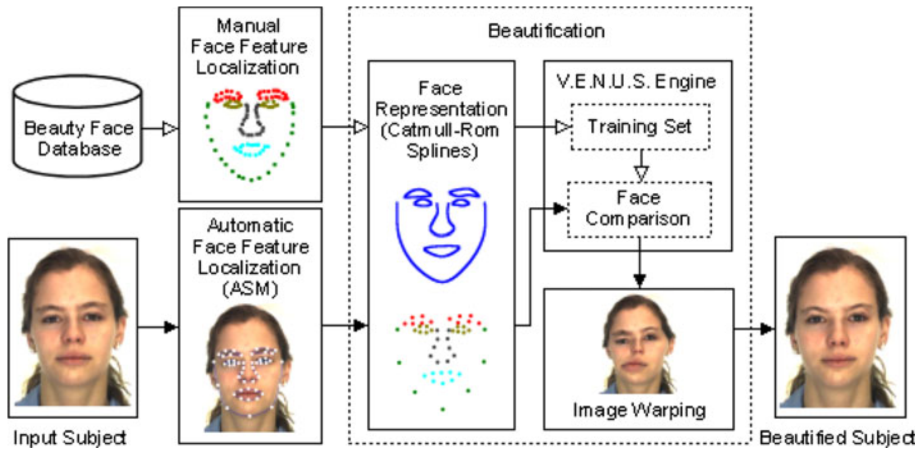

Fig. 2 The VENUS system

A quite simple way to approach this task is to define a set of explicit rules that can be applied to beautify a face. However, the definition and extraction of such rules is usually a complex process, and the resulting system is not flexible enough. VENUS adopts a machine learning approach to solve this particular task, and tries to learn from examples the optimal beauty concept for each subject. Examples are collected in a database of faces that are considered beautiful or attractive.

The definition of a sufficiently representative set of examples is not a trivial task. In fact, in order to avoid potential biases due to subjective criteria, the selection should be performed by different supervisors. However, when different judgments are collected, a way to merge these decisions should be considered, since the perception of beauty is not absolute. To solve those issues, in the current implementation, VENUS employs a set of images that depict celebrities, as actors, actresses, singers, or, generally, people who are universally considered beautiful.

Given the frontal-view pictures of those faces, their features were manually marked in order to extract the corresponding landmark-based representations. The flexibility of the model in defining corresponding points among different faces, its invariance to scaling and to in-plane head rotations, provide a stable and efficient method to compare different faces. Currently VENUS includes a training set composed of 30 portraits of beautiful men and 30 of beautiful women.

3.2 Face feature localization

An automatic face feature localization, based on ASMs [6], is embedded in VENUS. ASMs represent the shape of a face using a set of n points, so it is particularly suitable to work with the face representation model described in Sect. 2. ASMs exploit both a point distribution model (PDM), that describes the face shape variability estimated on a training set, and a luminance model (LM) for each landmark, used to move the face shape across the image in order to match the face position. The shape of a face is represented as a vector $\boldsymbol{f} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^{\mathrm{T}} \in \mathbb{R}^{2n}$. In order to estimate the PDM, a training set $\boldsymbol{F} = [\boldsymbol{f}_1|\boldsymbol{f}_2|\ldots|\boldsymbol{f}_m] \in \mathbb{R}^{2n,m}$, composed by m shapes is exploited. Such examples are aligned using the same coordinate frame and a uniform scaling factor, by means of Procrustes analysis [12]. During the alignment process, each shape is scaled so that $|\bar{\boldsymbol{f}}| = 1$, being $\bar{\boldsymbol{f}} = \frac{1}{m}\sum_{i=i}^{m} \boldsymbol{f}_i$, in order to guarantee the convergence of the Procrustes algorithm. After the alignment, the shapes form a distribution in $\mathbb{R}^{2n}$. Thus, the algorithm estimates a model $\boldsymbol{f} = M(\boldsymbol{b})$, where $\boldsymbol{b}$ is a set of parameters. An effective approach that allows us to define the PDM consists in applying PCA to the data. The mean $\bar{\boldsymbol{f}}$ and the covariance $\boldsymbol{S}$ of the distribution are computed. If $\Phi = [\Phi_1|\Phi_2|\cdots|\Phi_c] \in \mathbb{R}^{2n,c}$ and $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_c)^T \in \mathbb{R}^c$ collect the c leading eigenvectors and eigenvalues of $\boldsymbol{S}$, we can approximate a shape as

$$\boldsymbol{f} \approx \bar{\boldsymbol{f}} + \Phi\boldsymbol{b} \tag{2}$$

where $\boldsymbol{b} = \Phi'(\boldsymbol{f} - \bar{\boldsymbol{f}}) \in \mathbb{R}^c$ defines the parameters of the deformable model. The variance of b_i, the i_{th} entry of $\boldsymbol{b}$, across the training set, is given by the corresponding eigenvalue λ_i. Defining the variability range of b_i equal to $(-k\sqrt{\lambda_i}, k\sqrt{\lambda_i})$, being k a real constant, it is guaranteed that the generated shapes are similar to the examples in the training set. The shapes obtained using Eq. 2 are scaled to ensure the convergence of the alignment process. As a consequence, a transformation $T_{X_t,Y_t,s,\theta}(\bar{\boldsymbol{f}} + \Phi\boldsymbol{b})$, where X_t,Y_t represents a reference frame translation, s a scaling factor, and θ a rotation, must be applied to $\boldsymbol{f}$ to fit it on a given image. Hence, an instance $\boldsymbol{Y}$ of the PDM in the image frame can be created by defining the position, the scaling, the orientation and the set of parameters $\boldsymbol{b}$. The face feature localization is performed generating an initial shape $\boldsymbol{Y}^0 = (x_1^0, y_1^0, x_2^0, y_2^0, \ldots x_n^0, y_n^0)$ and then analyzing the image around each point (x_i^0, y_i^0) to find a better point position (x_i^1, y_i^1). Then,

$X_t, Y_t, s, \theta, \boldsymbol{b}$ are updated in order to match the new position $\boldsymbol{Y}^i = \left(x_1^1, y_1^1, x_2^1, y_2^1, \ldots x_n^1, y_n^1\right)$. This process is iteratively repeated until the displacement between the positions $\boldsymbol{Y}^t$ and $\boldsymbol{Y}^{t+1}$ is smaller than a certain threshold. The criterion used to find the optimal landmark position is based on a model of the local structure of the landmarks in the training set, i.e. estimating a luminance model for each landmark and looking for points that match the estimated LMs. Such models represent how the module of the luminance gradient varies nearby each landmark. Given the ith landmark and an image, w pixels on each side of the landmark along the normal to the shape boundary are considered and the $2w + 1$ samples of the module of the luminance gradient are collected in a vector $\boldsymbol{gl}_i$. The sampling is performed for each training image, obtaining a set of vectors for the given landmark. Assuming that such vectors are distributed as a multivariate Gaussian, the mean $\bar{\boldsymbol{gl}}$ and the covariance $\boldsymbol{S}_{lg}$ are estimated to represent a statistical model for the gradient of the luminance around the landmark. The Mahalanobis distance is used to compare an input vector $\boldsymbol{gl}_s$ with respect to the model, because this distance is inverse proportional to the log of the probability that $\boldsymbol{gl}_s$ is drawn from the distribution. During the face feature localization, z pixels ($z > w$) are sampled on each side of the landmark, along the normal to the boundary. Then, among the $2(z - w) + 1$ possible positions, the point that gives the lowest Mahalanobis distance is chosen. At a given iteration t, the search is repeated for each landmark, obtaining a suggested new position $\boldsymbol{Y}^{t+1}$ for the whole shape.

To set up VENUS, the PDM and the LMs were estimated using the AR [20] (125 images) and XM2VTS [21] (295 images) databases. Each face is represented using 49 landmarks, the other free parameters were chosen performing leave-one-out experiments and evaluating the results using the mean point distance metric. In particular, the shape variability parameter k was set to 2.0, the lengths of the luminance sampling w to 8 and z to 14, while the number of eigenvectors for the PDM estimation was selected such that $\sum_{i=1}^{c} \lambda_i \geq 0.97 \sum_{i=1}^{2n} \lambda_i$.

Finally, the choice of the initial shape position on the image reference frame is performed manually. The VENUS user interface allows to drag a line between the eyes of a depicted face in order to define the center of the reference frame (X_r, Y_r), its orientation θ and the scaling factor s, that is chosen equal to the inter-ocular distance. Then, the transformation $T(X_t, Y_t, s, \theta)$ is applied to the mean shape $\bar{\boldsymbol{f}}$ to generate the initial shape. The ASM algorithm generally performs quite well, in particular when the initialization process is carried out manually. At the end of the localization process, the facial features are displayed by the VENUS user interface. If some landmarks are quite far from their correct locations, the user can manually drag them to the desired position, before proceeding with the enhancement process.

3.3 Face comparison

The beauty concept exploited in the VENUS beautification process, derives directly from a comparison of faces in an appropriate feature space. VENUS represents each face as a vector $\boldsymbol{f} = (x_1, y_1, \ldots, x_{49}, y_{49})'$, whose pair of entries x_i, y_i collects the coordinates of the i_{th} landmark, expressed with respect to a reference frame centered in (X_r, Y_r) and rotated by an angle θ from the image frame.

In order to compare faces in this space a similarity measure, that represents the way humans perceive the similarity among different faces, is required. A simple solution to the face comparison is to assume that the face space is Euclidean. Note that, before comparing the coordinates of two faces, their landmarks must be aligned by performing a translation and a rotation that allow us to express the landmark coordinates in the same reference frame. If the processed face has an out-of-plane rotation, the comparison process is negatively biased, since according to the 2D landmark based representation, the face is considered particularly asymmetrical. Currently, this is a limitation of the system that can be overcome using a 3D face representation.

VENUS currently supports three different types of distances: Euclidean, weighted Euclidean, and Mahalanobis. The covariance matrix $\boldsymbol{C}$ used in the Mahalanobis distance is estimated from the beautiful face database (see Sect. 3.1), and it is accurately scaled in order to obtain $\det(\boldsymbol{C}) \neq 0$. In the case of the weighted Euclidean distance, the weights w_i are estimated from the same database, and correspond to the diagonal elements of $\boldsymbol{C}$, so $w_i = \frac{1}{\sigma_i^2}$, where σ_i is the standard deviation of the i_{th} component of the face vectors over the sample set.

3.4 Beautification

The "beautification process" of a human face picture can be defined as the enhancement of its beauty and attractiveness. VENUS tries to improve the attractiveness of the input image modifying the shape of the face features with respect to the faces in the beautiful face database; the only constraint to be satisfied is the recognizability of the original subject after the entire process.

The VENUS engine uses a K-nearest neighbors (KNN) classifier over the beauty face space in order to determine the shape transformations that will be applied to the input face. In details, the VENUS engine operates in three steps: face comparison, beauty target generation, and face merging.

The first step consists in the comparison of the input face vector against the vectors of the beautiful face database. Following the KNN algorithm, the K nearest faces $(\boldsymbol{f}_{n_1}, \ldots, \boldsymbol{f}_{n_K})$ and their distances to the input faces $(d_{n_1}, \cdots, d_{n_K})$ are considered. The next step consists in the generation of the beauty target face vector ($\boldsymbol{f}_{\mathrm{beauty}}$), that is obtained as the mean of the nearest faces weighted by the inverse of their distances from the input face,

$$\boldsymbol{f}_{\text{beauty}} = \frac{\sum_{j=1}^{K}\left(\frac{1}{d_{n_j}}\boldsymbol{f}_{n_j}\right)}{\sum_{j=1}^{K}\left(\frac{1}{d_{n_j}}\right)}.$$

The optimal value of K can be estimated experimentally. We found that a value of K between 3 and 6 seems to be a good choice. Using larger values makes the beauty target face too close to an average face, loosing some details and leading to a similar result for every input face (see Fig. 3a). The use of the inverse distances in the weighted average aims at avoiding this situation and it highlights the details of the closest templates. For example, comparing faces of Fig. 3b, c, we can observe that the nose of Fig. 3c is quite asymmetric and its eyebrows are particularly arched. In Fig. 3b those details are lost due to the unweighted average. Anyway, using larger values of K, this evident averaging effect still remains. On the other hand, the choice of a small number of neighbors leads to a beauty target that may preserve too much details of them, included noise introduced in the face feature extraction process.

The last step consists in merging the original input face with the beauty target. During this operation it should be considered that the final result must be a recognizable representation of the original subject, preserving his/her overall aspect, and slightly enhancing his/her facial features. The direct application of the beauty target face in substitution of the original one will result in something that is not related to the original face, even if it could be a qualitatively good result. VENUS merges the beauty target with the original face ($\boldsymbol{f}$) through a convex combination of them. The coefficient α of the combination determines the fraction of the original face that should be preserved. The resulting face is represented by the enhanced face vector ($\boldsymbol{f}_{\text{enhanced}}$),

$$\boldsymbol{f}_{\text{enhanced}} = \alpha\boldsymbol{f} + (1-\alpha)\boldsymbol{f}_{\text{beauty}} \quad (0 \le \alpha \le 1).$$

A value of α close to 0.5 seems to be the best choice in the general case. If a large number of nearest neighbors is chosen, then a smaller value of α is required to make the enhancement more effective. Using a value of K within the previously defined interval, α should range in $0.3 \le \alpha \le 0.5$. The use of different types of distance (see Sect. 3) leads to just slightly different results.

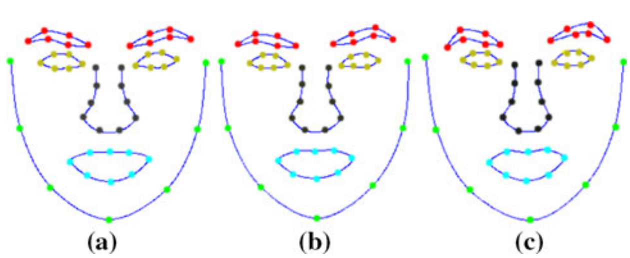

Fig. 3 Examples of beauty targets: **a** Generated with $K = 20$, **b**, **c** Generated with an unweighted and weighted mean in the case of $K = 3$

The final step in the VENUS beautification process is the transformation of the face portrait $\boldsymbol{f}$ according to the enhanced face vector $\boldsymbol{f}_{\text{enhanced}}$, by means of image warping. As stated in Sect. 3, the VENUS face representation is composed by 49 bidimensional points on the image plane. The facial image is modified using the pixels coordinates collected in $\boldsymbol{f}$ and $\boldsymbol{f}_{\text{enhanced}}$ that represent the source and the destination points of the image warping process.

VENUS implements a triangular mesh warping (TMW) [11] algorithm that establishes an affine mapping between corresponding triangles in the source and destination simplicial meshes. Those meshes are built over the vertex sets constituted by the source and destination points of the warping process. The TMW algorithm offers many useful properties, like the independence of the transformation of triangles that do not share any vertices, the simple computation of the warping function through barycentric coordinates, and the easy exploitation of the graphics processing unit (GPU). The first property makes it possible to enhance a single portion (triangle) of the face picture without altering the other parts of it (the parts that do not share any of the altered vertices). The main drawback of the TMW is the C^0 continuity between adjacent triangles. Anyway, since in the beautification process the alteration of the face is

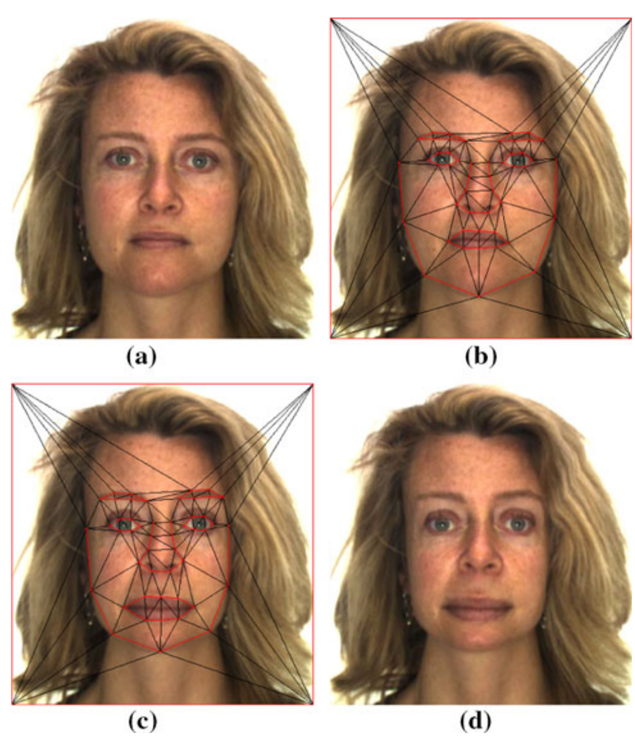

Fig. 4 Example of the triangular mesh warping process: **a** Original face, **b** Triangular mesh, **c** Mesh alteration, **d** Resulting (funny) face

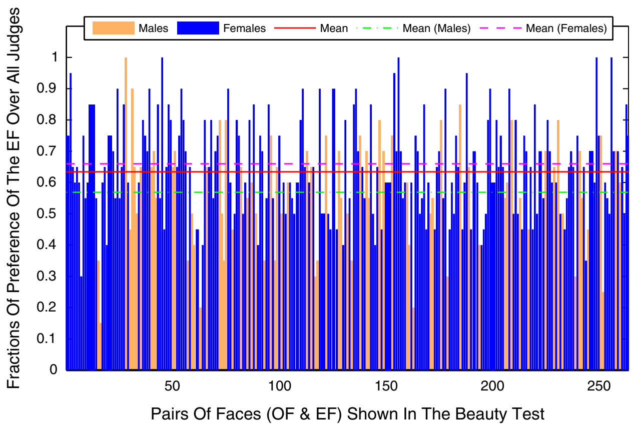

Fig. 5 Beauty test results. For each of the 265 pairs of faces, composed by the original face (*OF*) and the enhanced one (*EF*), the fraction of preferences for the beautified version over all judges is shown

Table 1 Beauty test results (*EFs* enhanced faces, *OFs* original faces)

	Men faces	Women faces	All faces
Mean preference Of EFs (%)	56.8	66.0	63.4
EFs majority voted (%)	61.3	78.9	73.9
OFs majority voted (%)	28.0	12.1	16.6
EFs voted by all judges (%)	1.3	2.1	1.8
OFs voted by all judges (%)	0.0	0.0	0.0
EFs–OFs equally voted (%)	10.6	8.9	9.4

The mean preference of EFs represents the mean percentage of preference for the beautified versions over all judges

minimal, this drawback is not evident. The use of a smoother warping function, such as thin plate splines (TPS) [4], would solve this issue at the cost of a higher complexity. However, the use of TPS does not ensure that a portion of the face will not be modified after the alteration of another one. VENUS realizes the warping process in the following steps. First, a Delaunay triangulation [23] is performed over the set of points of the input face, generating a simplicial mesh. The four points at the corners of the image are added to this set, since they constitute the vertices of the convex hull of the triangulation. The mesh is rendered by the GPU and textured using the original image. The entire vertex set is moved accordingly to the position of the enhanced face points, leaving unchanged the texture coordinates (see Fig. 4). The rendering and texturing phases employ the graphical capabilities of the GPU and the complete warping process is executed in real-time.

4 Results

The performances of VENUS have been accurately evaluated in practical contexts of use. We tested its capabilities on many portraits of various face databases, as AR or XM2VTS. The major drawback of the evaluation of beautified faces is that it is impossible to define a completely objective concept of beauty. For this reason, asking a single user to choose if the enhanced face is better than the original one is not sufficient. Moreover, the enhancements of pictures coming from benchmark face databases are easier to appreciate since they have been taken in controlled conditions. The heads are quite perfectly upright, the background is fixed and homogeneous, face expressions are completely neutral. Our goal was to test VENUS under the more realistic conditions, even when the subject is not in a perfectly neutral expression or with a frontal-posed head. For this reason, we acquired and enhanced a set of 265 face portraits (75 men and 190 women) during the Festival of Creativity. VENUS was set

Table 2 Beauty test results grouped by the sex of judges (*EFs* enhanced faces, *OFs* original faces)

		Men faces (%)	Women faces (%)	All faces (%)
Male judges	Mean preference of EFs (%)	57.7	68.9	65.7
	EFs majority voted (%)	70.6	87.8	83.0
	OFs majority voted (%)	29.3	12.1	16.9
	EFs voted by all judges (%)	1.3	3.1	2.6
	OFs voted by all judges (%)	0.0	0.0	0.0
	EFs–OFs equally voted (%)	0.0	0.0	0.0
Female judges	Mean preference of EFs (%)	54.1	57.2	56.3
	EFs majority voted (%)	62.6	62.6	62.6
	OFs majority voted (%)	37.3	37.3	37.3
	EFs voted by all judges (%)	4.0	10.0	8.3
	OFs voted by all judges (%)	1.3	3.1	2.6
	EFs–OFs equally voted (%)	0.0	0.0	0.0

The mean preference of EFs represents the mean percentage of preference for the beautified versions over all judges up to compare faces using the Euclidean distance. We did not take care of the light conditions or of the background. To analyze the quality of the produced results we used a web interface, called "VENUS beauty test", that shows a set of pages with the original and beautified picture of each subject, placed randomly to avoid biases, and gives the possibility to indicate the preferred one. We asked to 20 volunteers (judges) to express their preferences. In Fig. 5 the overall results of the VENUS beauty test are shown. For each of the 265 pairs of faces, the graph shows the percentage of preferences for the beautified version over all judges and in Table 1 some more detailed statistics are summarized.

The mean preference of beautified faces states that, for both men and women, the VENUS-enhanced face is preferred over the original one. In five cases (1 man and 4 women) the enhanced face has been chosen by all the judges, while none of the original face was preferred by all of them. Moreover, in the 73.9% of the cases, more than half of the judges preferred the beautified face to the original one.

In Table 2, the same statistics are divided into male (15) and female (5) judges. The mean preference expressed by women is quite constant among female and male subjects and it is lower with respect to the one expressed by men. This result seems to suggest that a different concept of beauty between different sexes exists. Finally, the evaluation of the results by means of majority voting allow us to appreciate the VENUS beautification performances.

Judges reported that for some face pairs they were really uncertain between the two proposed versions of the face. This can be noticed in the portion of the graph in Fig. 5 that is centered around 50% (0.5). Given a certain percentage threshold t, we define the portion of the graph that lies within 50% $\pm t$ as "uncertain region" . We can compute new statistics removing the pairs of faces whose judgment falls in such region. In Fig. 6, the resulting performances of VENUS are reported.

Even just removing pairs whose enhanced faces were preferred by exactly half of the judges ($t = 0$), the overall performances grows up to 81.6%, where for performances, in this context, we intend the percentage of beautified faces preferred by the majority of the judges. Performances keep growing with the size of the uncertain region. Obviously, at the same time, the number of evaluated pairs decreases, making the performance graph in Fig. 6 not particularly significant. A region size of about 20%, that corresponds to removing 94 subjects, appears to be consistent with judges' feedback. Removing larger regions leads to better performances (even 100.0%) but causes the removal of too many pairs, making the results less significative. Considering a region size of 20% ($t = 10$), the overall performances are 87.2%. In the case of female pairs, they are equal to 94.3%,

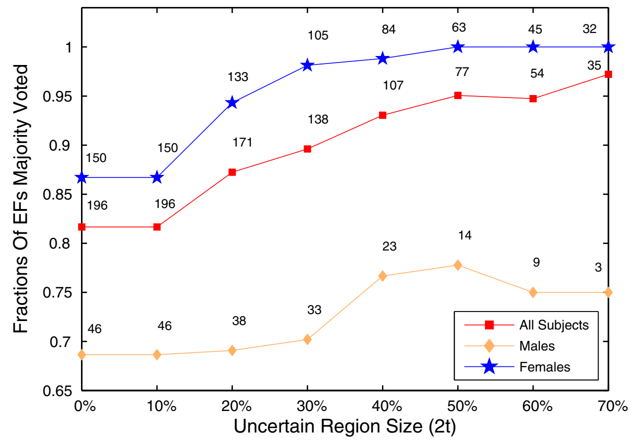

Fig. 6 Beauty test results after removing the "uncertain region". The *graph* shows the fraction of enhanced faces (*EFs*) preferred by the majority of the judges in function of the size of the region ($2t$). The number of subjects considered after the removal of the uncertain ones is reported nearby each point

Fig. 7 Some enhanced faces from the VENUS beauty test that were preferred by the majority of the judges. For each pair, the image on the *left* corresponds to the original one, while the VENUS result is reported on the *right*

showing that VENUS works extremely well with portraits of women. In the case of male pairs, performances reach the 69.0%.

Fig. 8 Some enhanced faces from the VENUS beauty test that were not preferred by the majority of the judges. For each pair, the image on the *left* corresponds to the original one, while the VENUS result is reported on the *right*

Fig. 9 Original faces and faces enhanced by VENUS from the AR database. For each pair, the image on the *left* corresponds to the original one, while the VENUS result is reported on the *right*

In Figs. 7 and 8 some beautifications that appear in the VENUS beauty test are reported, while in Fig. 9 some enhancements of face portraits of the AR face database are shown. More results can be found on VENUS website[2]. It is not very easy to define the reason why the enhanced faces reported in Fig. 8 were judged as less attractive with respect to their original version. As a matter of fact, the attractiveness of a face is a subjective perception. However, in some situations the warping phase produces a sort of artificial and unnatural aspect, that is judged negatively by the users, even if it is not properly an artifact. Probably, the drawback related to the warping phase can be partially overcome by increasing the number of subjects in the beautiful face database.

5 Conclusions

In this paper we described VENUS, a system for the automatic face enhancement. VENUS exploits a landmark-based representation of facial features, automatically localized by means of ASMs. An input image is compared against a database of faces that are universally considered particularly attractive. Then, it is enhanced moving the face features towards those defined by the K similar beautiful faces. The system performances have been evaluated by 20 volunteers, showing very promising results. VENUS has been developed with Java Technology (Java 6) and can be executed in any system that runs the Java Virtual Machine or can be easily embedded as an applet in a web site.

Future works include the extension of the standard ASMs algorithm in order to improve the face features localization, and the supervised learning of a similarity measure that allows to compare faces following the human perception of face similarity.

References

1. Alley TR, Cunningham MR (1991) Averaged faces are attractive, but very attractive faces are not average. Psychol Sci 2(2):123–125
2. Arakawa K, Nomoto K (2005) A system for beautifying face images using interactive evolutionary computing. In: Proceedings of the international symposium on intelligent signal processing and communication systems, pp 9–12
3. Avrithis Y, Xirouhakis Y, Kollias S (2001) Affine-invariant curve normalization for object shape representation, classification, and retrieval. Mach Vis Appl 13(2):80–94
4. Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE TPAMI 11(6):567–585
5. Catmull E, Rom R (1974) A class of local interpolating splines. Computer Aided Geometric Design, pp 317–326
6. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models—their training and application. Comput Vis Image Underst 61(1):38–59
7. Davies RH (2002) Learning shape: optimal models for analysing natural variability. PhD thesis, Department of Imaging Science, University of Manchester
8. Eisenthal Y, Dror G, Ruppin E (2006) Facial attractiveness: beauty and the machine. Neural Comput 18(1):119–142
9. Farkas LG (1987) Linear proportions in above-and below-average women's faces. Anthropometric facial proportions in medicine, 3rd edn. Charles C. Thomas, Springfield, pp 119–129
10. Fink B, Grammer K, Thornhill R (2001) Human (*Homo sapiens*) facial attractiveness in relation to skin texture and color. J Comp Psychol 115(1):92–99

2 http://venus.dii.unisi.it

11. Gomes J, Costa B, Darsa L, Velho L (1998) Warping and morphing of graphics objects. Morgan Kaufmann, San Francisco
12. Goodal C (1991) Procrustes methods in the statistical analysis of shapes. J Roy Stat Soc 53(2):285–339
13. Gunes H, Piccardi M (2006) Assessing facial beauty through proportion analysis by image processing and supervised learning. Int J Hum Comput Stud 64(12):1184–1199
14. Jefferson Y (2004) Facial beauty—establishing a universal standard. Int J Orthod 15(1):9–22
15. Langlois JH, Roggman LA (1990) Attractive faces are only average. Psychol Sci 1(2):115–121
16. Leyvand T, Cohen-Or D, Dror G, Lischinski D (2006) Digital face beautification. In: SIGGRAPH '06, ACM SIGGRAPH 2006 Sketches. ACM, New York, p 169
17. Liu H, Yan J, Li Z, Zhang H (2007) Portrait beautification: a fast and robust approach. Image Vis Comput 25:1404–1413
18. Liu Z, Zhang C, Zhang Z (2007) Learning-based perceptual image quality improvement for video conferencing. In: Proceedings of IEEE international conference on multimedia and expo, 2007, pp 1035–1038
19. Maggini M, Melacci S, Sarti L (2007) Representation of facial features by Catmull–Rom splines. In: Proceedings of international conference on computer analysis of images and patterns. Springer, Berlin, pp 408–413
20. Martinez AM, Benavente R (1998) The AR face database. CVC Technical Report 24
21. Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999) XM2VTSDB: the extended M2VTS database. In: International conference on audio and video-based biometric Person authentication, pp 72–77
22. Michiels G, Sather AH (1994) Determinants of facial attractiveness in a sample of white women. Int J Adult Orthod Orthognath Surg 9(2):95–103
23. Su P, Drysdale RLS (1995) A comparison of sequential Delaunay triangulation algorithms. In: Proceedings of the 11th symposium on computational geometry, pp 61–70
24. Swaddle JP, Cuthill IC (1995) Asymmetry and human facial attractiveness: symmetry may not always be beautiful. Proc Biol Sci 261(1360):111–116
25. Wang H, Kearney J, Atkinson K (2002) Arc-length parameterized spline curves for real-time simulation. In: International conference on curves and surfaces, pp 387–396

Author Biographies

Stefano Melacci received the Laurea degree in Computer Engineering (cum laude) from the University of Siena, Italy, in 2006. In the same year he started a Ph.D. program in Information Engineering and currently he is a Ph.D. student at the Department of Information Engineering, University of Siena. His research interests include machine learning and pattern recognition, mainly focused on learning of similarity measures, regularization theory and applications to human face analysis. He is an active reviewer for many international conferences and journals.

Lorenzo Sarti received the Laurea degree in Computer Science from the University of Firenze, Italy, in 2001, and the Ph.D. degree in Computer Engineering from the University of Siena, Italy, in May 2005. Currently, he is a Research Associate in Information Engineering at the University of Siena, Italy. His main research interests are in pattern recognition in structured domains, machine learning techniques applied to image understanding and visual information retrieval. He authored 26 publications in international journals and conferences and has been involved in some research projects focused on neural networks and pattern recognition at the University of Siena.

Marco Maggini received the Laurea degree cum laude in electronic engineering from the University of Firenze in February 1991 and the Ph.D. in computer science and control systems in 1995 from the University of Firenze. In February 1996, he became an assistant professor of computer engineering at the School of Engineering of the University of Siena, where, since March 2001, he has been an associate professor. His main research interests are machine learning, neural networks, human–machine interaction, technologies for distributing and searching information on the Internet, and non-structured databases. He has been collaborating with the NEC Research Institute, Princeton, New Jersey, on parallel processing, neural networks, and financial time series prediction. He is an associate editor of the *ACM Transactions on Internet Technology*. He has been a guest editor of a special issue of the *ACM TOIT* on machine learning and the Internet. He contributed to the organization of international and national scientific events. He is member of the IAPR-IC and the IEEE Computer Society.

Marco Gori received the Ph.D. degree in 1990 from University of Bologna, Italy. From October 1988 to June 1989 he was a visiting student at the School of Computer Science (McGill University, Montreal). In 1992, he became an Associate Professor of Computer Science at University of Firenze and, in November 1995, he joined the University of Siena, where he is currently full professor of computer science. His main interests are in machine learning, with applications to pattern recognition, Web mining, and game playing. He is especially interested in the formulation of relational machine learning schemes in the continuum setting. Prof. Gori serves (has served) as an Associate Editor of a number of technical journals related to his areas of expertise, including *IEEE Transaction on Neural Networks*, *Pattern Recognition*, *Neural Networks*, *Neurocomputing*, *Pattern Analysis and Application*, the *International Journal of Document Analysis and Recognition*, and the *International Journal on Pattern Recognition and Artificial Intelligence*. He has been the recipient of best paper awards and keynote speakers in a number of international conferences. He is the Chairman of the Italian Chapter of the IEEE Computational Intelligence Society, a fellow of the ECCAI and the IEEE.