Generating Discriminating Cartoon Faces Using Interacting Snakes

Rein-Lien Hsu, *Member*, *IEEE*, and Anil K. Jain, *Fellow*, *IEEE*

Abstract—As a computational bridge between the high-level a priori knowledge of object shape and the low-level image data, active contours (or snakes) are useful models for the extraction of deformable objects. We propose an approach for manipulating multiple snakes iteratively, called *interacting snakes*, that minimizes the attraction energy functionals on both contours and enclosed regions of individual snakes and the repulsion energy functionals among multiple snakes that interact with each other. We implement the interacting snakes through explicit curve (parametric active contours) representation in the domain of face recognition. We represent human faces semantically via facial components such as eyes, mouth, face outline, and the hair outline. Each facial component is encoded by a closed (or open) snake that is drawn from a 3D generic face model. A collection of semantic facial components form a hypergraph, called *semantic face graph*, which employs interacting snakes to align the general facial topology onto the sensed face images. Experimental results show that a successful interaction among multiple snakes associated with facial components makes the semantic face graph a useful model for face representation, including cartoon faces and caricatures, and recognition.

Index Terms—Active contours, snakes, gradient vector field, face recognition, semantic face graph, face modeling, face alignment, cartoon faces, caricatures.

1 Introduction

Over the past decade, face recognition has received substantial attention from researchers in pattern recognition, computer vision, and cognitive psychology communities (see the survey in [25]). This common interest is motivated by challenges in designing machine vision systems that will match our remarkable ability to recognize people based on facial features, by the increased attention being devoted to security applications, and by the growing need of automatic image and video archival based on human faces. The main challenge in face recognition is the presence of a large intraclass variability in human face images due to 3D head pose, lighting, facial expression, facial hair, and aging, and rather small intersubject variations (due to similarity of individual appearances).

Face recognition algorithms can be classified as pose-dependent and pose-invariant. In pose-dependent algorithms, a face is represented by a small number of 2D images (appearances) at different poses, a set of viewer-centered images. On the other hand, in pose-invariant approaches, a face is represented by a 3D model, an object-centered representation. The pose-dependent algorithms can be further classified into three major groups: 1) the geometry-based approach uses the configuration of geometrical features of the face [7], 2) the appearance-based approach uses *holistic* texture features [20], and 3) the hybrid approach combines facial geometry and local appearance information [22], [16]. The geometry-based methods suffer from an insufficient number of facial landmarks that often cannot be detected accurately; the appearance-based techniques are unable to tolerate variations in head pose, facial expression, and illumination. The pose-invariant algorithms use 3D face models that are promising to overcome the above-mentioned variations, although it is difficult to align 3D face structure with 2D images and is cost-sensitive to acquire 3D face shape. Therefore, grouping low-level features (such as locations of feature landmarks, texture, and 3D head surface [10]) into a meaningful semantic entity (e.g., nose, mouth, and eyes) has become an attractive approach to face recognition.

Modeling facial components at the semantic level can help us to understand how the individual components contribute to face recognition. People can easily identify faces in caricatures (see Figs. 1a, 1b, and 1c) that exaggerate some of the salient facial components. Caricatures reveal that there are certain facial features which are salient for each individual and that a relatively easier identification of faces can occur by emphasizing distinctive facial components and their configuration. Further, two cartoon faces, as shown in Figs. 1d and 1e, reveal that line drawings and color characteristics (shades) of facial components provide sufficient information for humans to recognize the faces in cartoon movies. People can still recognize faces without the use of shading information, which is rather unstable under different lighting conditions. However, very little work has been done in face recognition based on facial sketches [21] and (computer-generated [3]) caricatures [13], [18].

We propose a *semantic* and potentially pose-invariant approach for face recognition based on a generic 3D face model. From a 3D face model, we can derive 2D *semantic face graphs* for identifying faces at a semantic level. Each facial component is modeled by its open (or closed) boundary using an active contour (snake). Research on active contours focuses

- *R.-L. Hsu is with Research Group, Identix Inc., Jersey City, NJ 07302. E-mail: Vincent.Hsu@identix.com.*
- *A.K. Jain is with the Department of Computer Science and Engineering, Michigan State University, MI 48824. E-mail: jain@cse.msu.edu.*

Manuscript received 1 May 2002; revised 23 Feb. 2003; accepted 3 June 2003. Recommended for acceptance by M.A.T. Figueiredo, E.R. Hancock, M. Pelillo, and J. Zerubia.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118717.

0162-8828/03/$17.00 © 2003 IEEE Published by the IEEE Computer Society

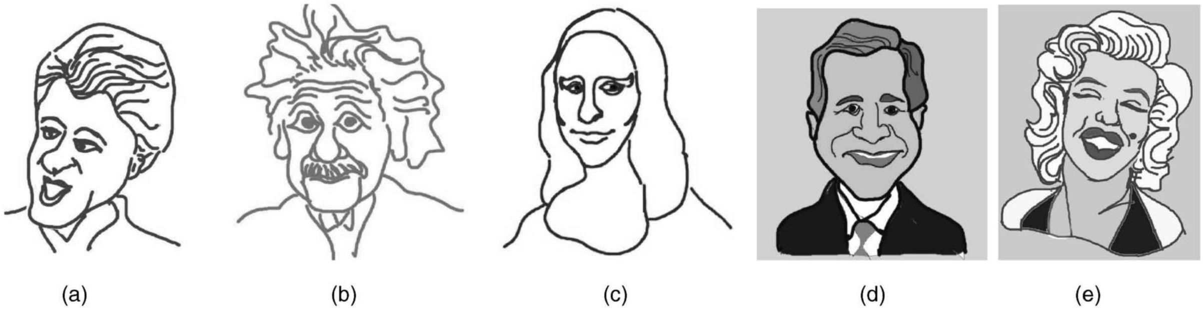

Fig. 1. Caricatures of (a) Bill Clinton, (b) Albert Einstein, and (c) the Mona Lisa. Cartoon faces of (d) George W. Bush and (e) Marilyn Monroe. (All these pictures were illustrated by R.-L. Hsu.)

on issues related to representation (e.g., parametric curves, splines, Fourier series, and implicit level-set functions), initialization, energy functionals, implementations (e.g., classical finite difference models, dynamic programming [2], and Fourier spectral methods), convergence rates and conditions, and their relationship to statistical theory [14] (e.g., the Bayesian estimation). Classical snakes [12] are represented by parametric curves and are deformed via finite difference methods based on edge energies. Different types of edge energies including image gradients, gradient vector flows [23], distance maps, and balloon force have been proposed. Snakes implicitly combined with level-set methods based on the curve evolution theory, called geodesic active contours (GAC) [8], are used to extract unknown geometric topology of close curves. Besides the edge energy, region energy has been introduced to improve the segmentation results for homogeneous objects in both the parametric and the GAC approaches (e.g., region and edge [15], GAC without edge [5], statistical region snake [6], region competition [26], and active region model [11]). Multiple active contours [1], [4] have been proposed to extract/partition multiple *homogeneous* regions that do not overlap with each other in an image.

We utilize face detection results (face and eye locations) to initialize multiple snakes that represent the complete face graph and interact with each other to extract an aligned face graph (called a cartoon face) for face matching. Since, facial components usually overlap, e.g., eyes are inside the face outline, we introduce a repulsion force among multiple parametric contours for preserving facial topology. We propose an approach for manipulating multiple snakes iteratively, called *interacting snakes*, that minimizes the attraction energy functionals on both contours and enclosed regions of individual snakes and the repulsion energy functionals among multiple snakes. We have implemented the interacting snakes through explicit curve (i.e., parametric active contours) representations for face alignment. Once the semantic face graph has been aligned to face images, we generate facial caricatures, and derive component weights for face matching, based on distinctiveness and visibility of individual components. Face matching is performed at a semantic level in a feature space spanned by Fourier descriptors of facial components.

2 Semantic Face Graph

A semantic face graph provides a high-level description of the human face. A projected graph in frontal view is shown in Figs. 1a, 1b, and 1c. A node of the graph represents a semantic facial component (e.g., eyes and mouth), each of which is constructed from a subset of vertices in the 3D generic face model and is enclosed by parametric curves. A semantic graph is represented in a 3D space and is compared with other such graphs in a 2D projection space. Therefore, the 2D appearance of the semantic graph looks different at different viewpoints due to the effect of perspective projection of the facial surface. We adopt Waters' animation model [19] as the generic face model because it contains all the internal facial components, face outline, and muscle models for mimicking facial expressions. However, Waters' model does not include some of the crucial external facial features. Hence, we have created external facial components such as the ear and the hair contours for the frontal view of Waters' model. We hierarchically decompose the vertices of the mesh model into three levels: 1) vertices at the boundaries of facial components, 2) vertices constructing facial components, and 3) vertices belonging to facial skin regions. The vertices at the top level are labeled with facial components such as the face outline, eyebrows, eyes, nose, and mouth using curves (Fig. 2d). The coordinates of a component boundary can also be represented by parametric curves, i.e., $c(s) = (x(s), y(s))$, where $s \in [0, 1]$, which is a snake for explicit curve deformation or for generating level-set functions for implicit curve evolution.

3 Coarse Alignment of Semantic Face Graph

Face modeling (alignment) is one of the three major modules (others being face detection and recognition) in our face recognition system. It is decomposed into coarse and fine alignment (described in Section 4) submodules. In the coarse alignment, a semantic face graph adapts to a face image through the *global and local* rigid 3D geometric transformation (scaling, rotation, and translation), based on the detected locations of face and facial components (see Figs. 3a, 3b, 3c, and 3d for detection results using the algorithm in [9]). Currently, we assume that all of the internal and external facial components of a face image are visible to the modeling module. We further employ the edges and color characteristics of facial components to locally refine the rotation, translation, and scaling

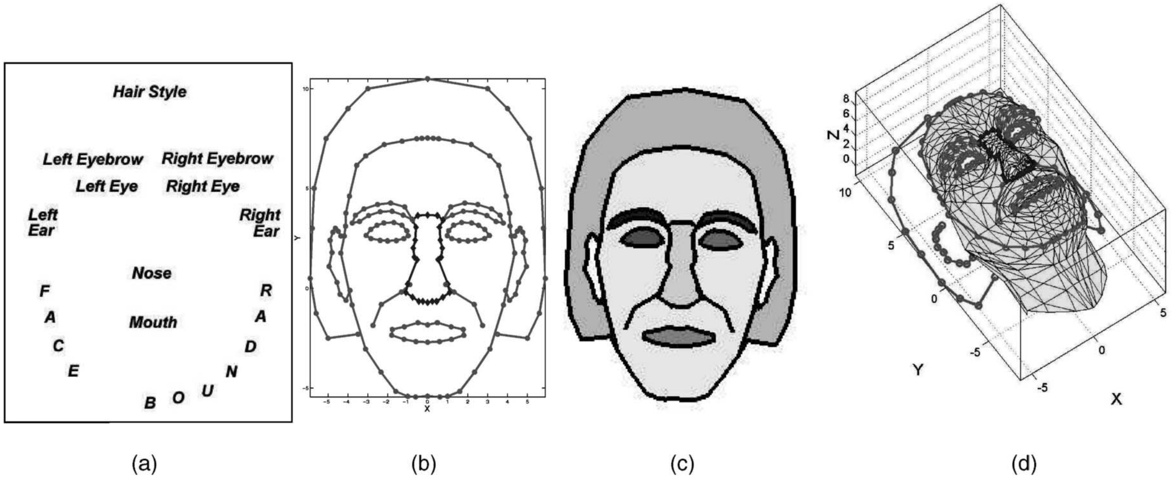

Fig. 2. Semantic face graph is shown in frontal view, whose nodes are (a) indicated by text, (b) depicted by polynomial curves, (c) filled with different shades, and (d) overlaid on a 3D generic face model in side view.

parameters for individual components. This parameter refinement is achieved by maximizing a *semantic facial score* (SFS) through a small amount of perturbation of the parameters. The semantic facial score of a component set T on a face image $I(u,v)$, SFS_T, is defined by a priori weights on facial components and component matching scores as follows:

$$SFS_T = \frac{\sum_{i=0}^{N-1} wt(i) \cdot MS(i)}{\sum_{i=0}^{N-1} wt(i)} - \rho \cdot SD(MS(i)), \qquad (1)$$

where N is the number of semantic components in T, $wt(i)$, and $MS(i)$ are, respectively, the a priori weight and the matching score of component i, ρ is a constant used to penalize the components with high standard deviations of the matching scores, and $SD(x)$ stands for the standard deviation of x. The matching score for the ith facial component is computed based on the coherence of the boundary and the coherence of color content (represented by a component map) by

$$\begin{aligned} MS(i) = &\frac{1}{M_i}\sum_{j=0}^{M_i-1}\left(\frac{1}{A_i}\sum_{k=0}^{A_i-1} e(u_k,v_k)\right) \\ &\cdot \frac{\left|cos(\theta_i^G(u_j,v_j) - \theta(u_j,v_j))\right| + f(u_j,v_j)}{2}, \end{aligned} \qquad (2)$$

where M_i and A_i are, respectively, the number of pixels along the curve of component i and number of pixels covered by the component i, θ_i^G and θ are the normal directions of component curve i in a semantic graph G and the gradient orientation of image I, f is the edge magnitude of the image I, and $e(u_k,v_k)$ is the facial component map of the image I at pixel k. The gradient magnitude, gradient orientation, eye map (See [9] and (7)), and coarse alignment results for the subject in Fig. 3c are shown in Fig. 4.

4 Fine Alignment of Semantic Face Graph

Fine alignment employs multiple (closed or open) snakes to locally deform a semantic face graph through a repulsion energy from a general facial topology to a sensed face image iteratively. We have studied two competing implementations of active contours for deforming interacting snakes: 1) explicit (or parametric) and 2) implicit contour representations. The explicit contour representation has the advantage of maintaining the geometric topology, while the implicit contour representation requires topological constraints on implicit functions. We implement interacting snakes via the parametric approach, because it can easily constrain the facial topology.

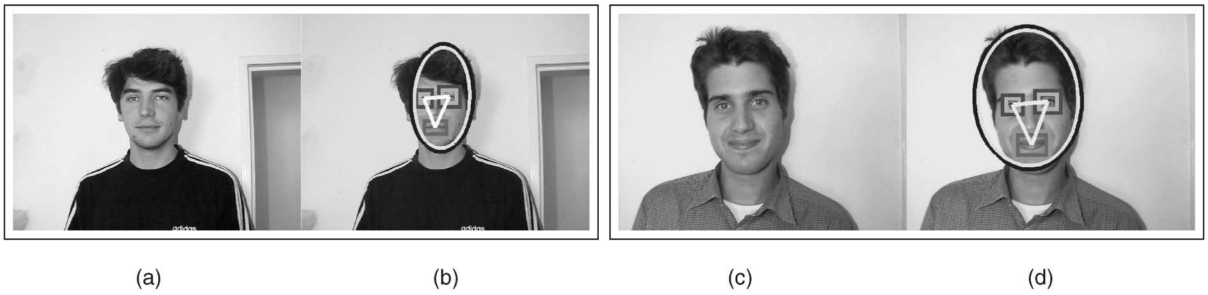

Fig. 3. Face detection results: (a) and (c) are input face images of size 640×480 from the MPEG7 content set (See [12]). (b) and (d) are detected faces, each of which is described by an oval and a triangle.

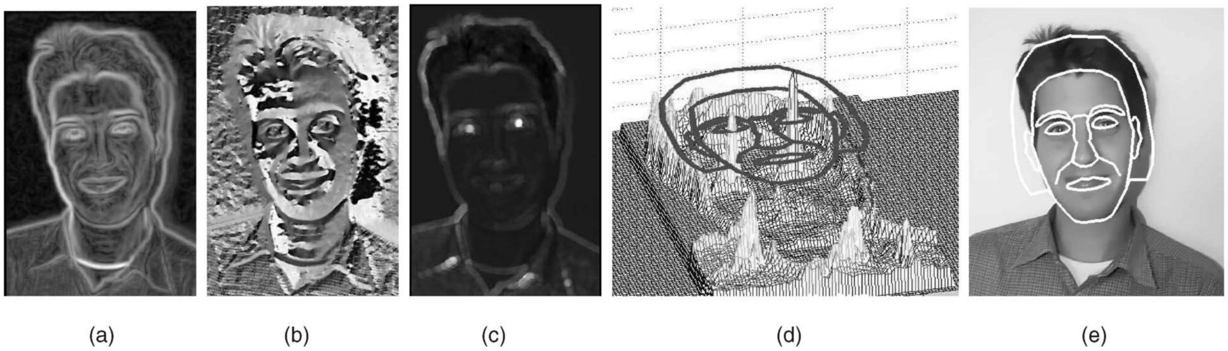

Fig. 4. Boundary map and eye component map for coarse alignment: (a) and (b) are gradient magnitude and orientation, respectively, obtained from multiscale Gaussian-blurred edge response, (c) an eye map extracted from the face image shown in Fig. 3c, (d) a semantic face graph overlaid on a 3D plot of the eye map, and (e) image overlaid with a coarsely aligned face graph.

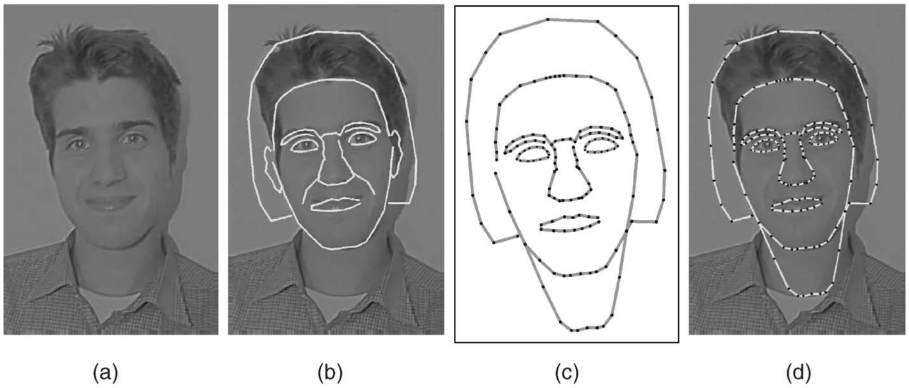

Fig. 5. Initialization of interacting snakes: (a) Face candidate extracted from the face image shown in Fig. 3a. (b) Coarsely aligned semantic face graph overlaid on the face candidate. (c) Initial configuration of interacting snakes. (d) The interacting snakes shown in (c) overlaid on the face candidate.

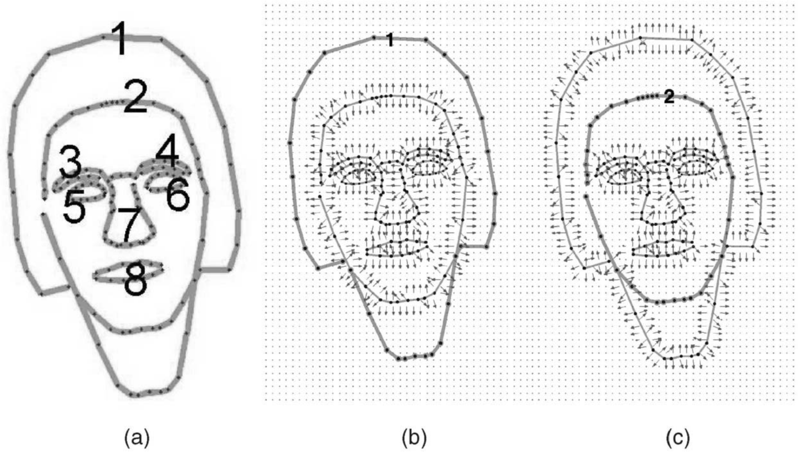

Fig. 6. Repulsion force: (a) interacting snakes with index numbers marked, (b) the repulsion force computed for the hair outline, and (c) the repulsion force computed for the face outline.

4.1 Interacting Snakes and Energy Functional

The initial configuration of interacting snakes is obtained from the coarsely-aligned semantic face graph, and is shown in Fig. 5d. Currently, there are eight snakes in our model that interact with each other. These snakes describe the hair outline, face outline, eyebrows, eyes, nose, and mouth of a face; they are denoted as $V(s) = \bigcup_{j=1}^{N} \{v_i(s)\}$, where N $(= 8)$ is the number of snakes, and $v_i(s)$ is the ith snake with parameter $s \in [0, 1]$. The energy functional used by interacting snakes is described in (3).

$$E_{isnake} = \sum_{i=1}^{N} \left[\int_0^1 \underbrace{E_{internal}(v_i(s)) + E_{repulsion}(v_i(s))}_{E_{prior}} + \underbrace{E_{attraction}(v_i(s))}_{E_{observation}} \, ds \right], \tag{3}$$

where i is the index of the interacting snake. The first two terms in (3) are based on the prior knowledge of snake's shape and snake's configuration (i.e., facial topology), while the third term is based on the sensed image (i.e., observed pixel

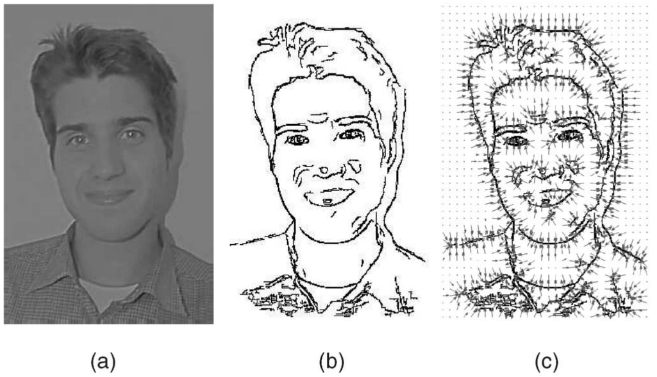

Fig. 7. Gradient vector field: (a) face region of interest extracted from a 640×480 image, (b) thresholded gradient map based on the population of edge pixels shown as dark pixels, and (c) gradient vector field.

values). The internal energy consists of smoothness and stiffness energies of a contour, while the repulsion energy is constructed among multiple snakes. The attraction energy is drawn from the image around the contours themselves and their enclosed regions. From calculus of variations, we know that interacting snakes which minimize the energy function in (3) must satisfy the following Euler-Lagrange equation:

$$\sum_{i=1}^{N}\left[\underbrace{\alpha v_i''(s) - \beta v_i^{(4)}(s)}_{\text{Internal Force}}\underbrace{-\nabla E_{repulsion}(v_i(s))}_{\text{Repulsion Force}}\underbrace{-\nabla E_{attraction}(v_i(s))}_{\text{Attraction Force}}\right] = 0, \tag{4}$$

where α and β are coefficients for adjusting the second and the fourth order derivatives of a contour, respectively. Repulsion force field is constructed based on the gradients of distance map among the interacting snakes as follows:

$$-\nabla E_{repulsion}(v_i(s)) = \lambda \cdot \nabla\left(1 - e^{C\cdot EDT\left(\bigcup_{j=1,j\neq i}^{N} v_j(s)\right)}\right), \tag{5}$$

where repulsion weight $\lambda = 0.81$, control factor $C = 3.9$, and EDT is the Euclidean Distance Transform. Fig. 6 shows the repulsion force fields for the hair outline and the face outline. The use of the repulsion force can prevent different active contours from converging to the same location of minimum energy.

The attraction force field consists of two kinds of fields in (6): one is obtained from edge strength, called gradient vector field (GVF) [23], and the other from a region pressure field (RPF) [11].

$$\begin{aligned} -\nabla E_{attration}(v_i(s)) &= GVF + RPF \\ &= \kappa \cdot \vec{\mathbf{V}}(v_i(s)) + \rho \cdot \vec{\mathbf{N}}(v_i(s)) \\ &\quad \cdot \left(1 - \frac{|E_i^{comp}(v_i(s)) - \mu|}{k\sigma}\right), \end{aligned} \tag{6}$$

where $\vec{\mathbf{V}}$ is the gradient vector flow field with flow weight $\kappa = 0.9$; $\vec{\mathbf{N}}(v_i(s))$ is the normal vector to the $i\text{th}$ contour $v_i(s)$ with pressure weight $\rho = 0.25$; E_i^{comp} is the component energy of the $i\text{th}$ component; μ, σ are the mean and the standard deviation of region energy over a seed region of the $i\text{th}$ component; $k(= 20)$ is a constant that constrains the energy variation of a component. The advantage of using GVF for snake deformation is that its range of influence is larger than that obtained from gradients and can attract snakes to a concave shape. A GVF is constructed from an edge map by an iterative process. However, the construction of GVF is very sensitive to noise in the edge map; hence, it requires a clean edge map as an input. Therefore, we compute a GVF by using three edge maps obtained from luma and chroma components of a color image, and by choosing as the edge pixels the top $p\% (= 15\%)$ of edge pixel population over a face region, as shown in Fig. 7a. Fig. 7b is the edge map for constructing the GVF that is shown in Fig. 7c). The region pressure field is available only for a homogeneous region in the image. However, we can construct component energy maps that reveal the color property of facial components such as eyes with bright-and-dark pixels and mouth with red lips. Then, a region pressure field can be calculated based on the component energy map and on the mean and standard deviation of the energy over seed regions (note that we know the approximate locations of eyes and mouth). Let us denote the color components in the RGB space as (R, G, B), and those in YCbCr space as (Y, Cb, Cr). An eye component energy for a color image is computed as follows:

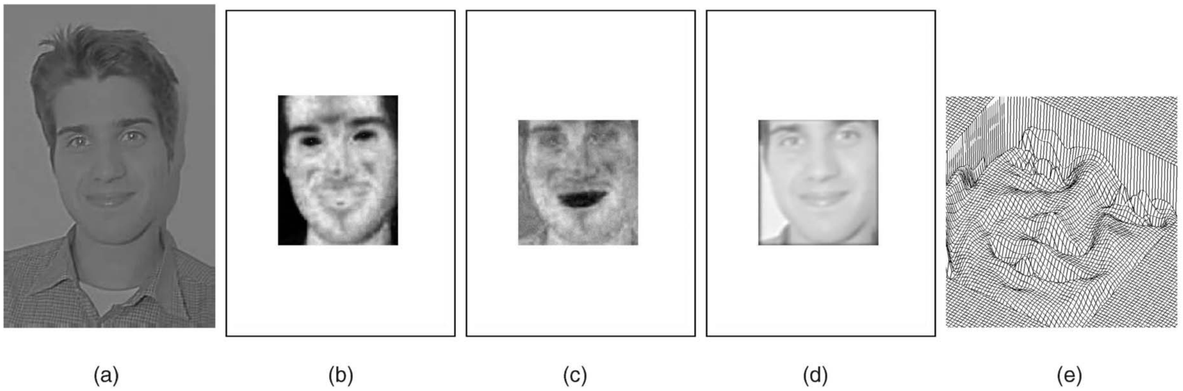

Fig. 8. Component energy (darker pixels have stronger energy): (a) face region of interest, (b) eye component energy, (c) mouth component energy, (d) nose boundary energy, and (e) magnified nose boundary energy shown as a 3D mesh surface.

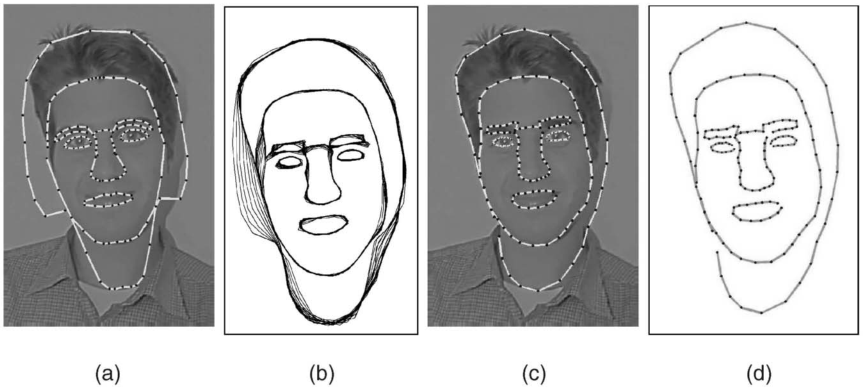

Fig. 9. Fine alignment: (a) interacting snakes overlaid on a face candidate, (b) snake deformation shown with 16 epochs (five iterations per epoch), (c) aligned snakes (currently eight snakes for hair outline, face border, eyebrows, eyes, nose, and mouth are interacting) overlaid on the face candidate, and (d) aligned snakes shown alone.

$$E_{eye}^{comp} = E_{msat} + E_{csh} + E_{cdif}, \tag{7}$$

$$E_{msat} = \left[\left(\left(R - \frac{K}{3} \right)^2 + \left(G - \frac{K}{3} \right)^2 + \left(B - \frac{K}{3} \right)^2 - \frac{(R + G + B - K)^2}{3} \right)^{0.5} \right], \tag{8}$$

$$E_{csh} = [[Cr - K/2]^2 - [Cb - K/2]^2], \tag{9}$$

$$E_{cdif} = [[Cr] - [Cb]], \tag{10}$$

where E_{msat} is the modified saturation (that is the distance in a plane between a point (R, G, B) and $(K/3, K/3, K/3)$), where $R + G + B = K$, E_{csh} is chroma shift, E_{cdif} is chroma difference, $K = 256$ is the number of grayscales for each color component, and $[x]$ indicates a function that normalizes x into the interval $[0, 1]$. The eye component energy for the subject in Fig. 8a is shown in Fig. 8b. The mouth component energy is computed as $E_{mouth}^{comp} = [-[Cb] - [Cr]]$. Fig. 8c shows an example of mouth energy. For the nose component, its GVF is usually weak, and, therefore, it is difficult to construct an energy map for nose. Hence, for the nose, we utilize a shape-from-shading (SFS) algorithm [24] to generate a boundary energy for augmenting the GVF for the nose component. Figs. 8d and 8e show the nose boundary energy as a 2D grayscale image and a 3D mesh plot, respectively.

4.2 Parametric Active Contours

Once we obtain the attraction force, we can make use of the implicit finite differential method [12], [23] and the iteratively updated repulsion force to deform the snakes. The repulsion force is computed and merged with the attraction force in each iteration via the weight λ. The stopping criteria is based on limits of iterative movement of each snake. Fig. 9a shows the initial interacting snakes, Fig. 9b shows snake deformation, and Figs. 9c and 9d show finely aligned snakes.

5 Semantic Face Matching and Facial Caricatures

For face matching, we construct a face descriptor in spatial frequency domain based on the Fourier transform of a semantic face graph. Let the semantic graph projected on a 2D image represented by the set $\mathbf{T}$ be $\mathbf{G}$. The coordinates of component boundary of $\mathbf{G}$ can be represented by a pair of sequences $x_i(n)$ and $y_i(n)$, where $n = 0, 1, \ldots, N_i - 1$ and $i = 1, \ldots, M$, for component i with N_i vertices. The 1D Fourier transform, $a_i(k)$, of the complex signal $u_i(n) = x_i(n) + jy_i(n)$ (where $j = \sqrt{-1}$) is computed by

$$a_i(k) = \mathcal{F}\{u_i(n)\} = \sum_{n=0}^{N_i - 1} u_i(n) \cdot e^{-j2\pi kn/N_i}, \tag{11}$$

for facial component i with a close boundary such as eyes and mouth, and with end-vertex padding for components with open boundary such as ears and hair components. The advantage of using semantic graph descriptors for face matching is that these descriptors can seamlessly encode geometric relationships (scaling, rotation, translation, and shearing) among facial components in a compact format. The reconstruction of semantic face graphs from semantic graph descriptors is obtained by

$$\tilde{u}_i(n) = \mathcal{F}^{-1}\{a_i(k)\} = \sum_{k=0}^{L_i - 1} a_i(k) \cdot e^{j2\pi kn/N_i}, \tag{12}$$

where L_i $(< N_i)$ is the number of frequency components used for the $i\mathrm{th}$ face component.

5.1 Component Weights and Matching Cost

After the two phases of face alignment, we can automatically derive a weight (called *semantic component weight*) for each facial component i for a subject P with N_p training face images by

$$scw^P(i) = \begin{cases} 1 + e^{-2\sigma_d^2(i)/d^2(i)} & N_p > 1, \\ 1 + e^{-1/d^2(i)} & N_p = 1, \end{cases} \tag{13}$$

$$d(i) = \frac{1}{N_P} \sum_{k=1}^{N_P} SFD_i(\mathbf{G_0}, \mathbf{G_{P_k}}) \cdot MS^{P_k}(i), \tag{14}$$

$$\sigma_d(i) = SD_k\big[SFD_i(\mathbf{G_0}, \mathbf{G_{P_k}}) \cdot MS^{P_k}(i)\big], \tag{15}$$

where SFD means semantic facial distance, MS is the matching score, SD stands for standard deviation, and $\mathbf{G_0}$ and $\mathbf{G_{P_k}}$ are the coarsely aligned and finely deformed semantic face graphs, respectively. The semantic component weights take values between 1 and 2. The semantic facial distance of facial component i between two graphs is defined as follows:

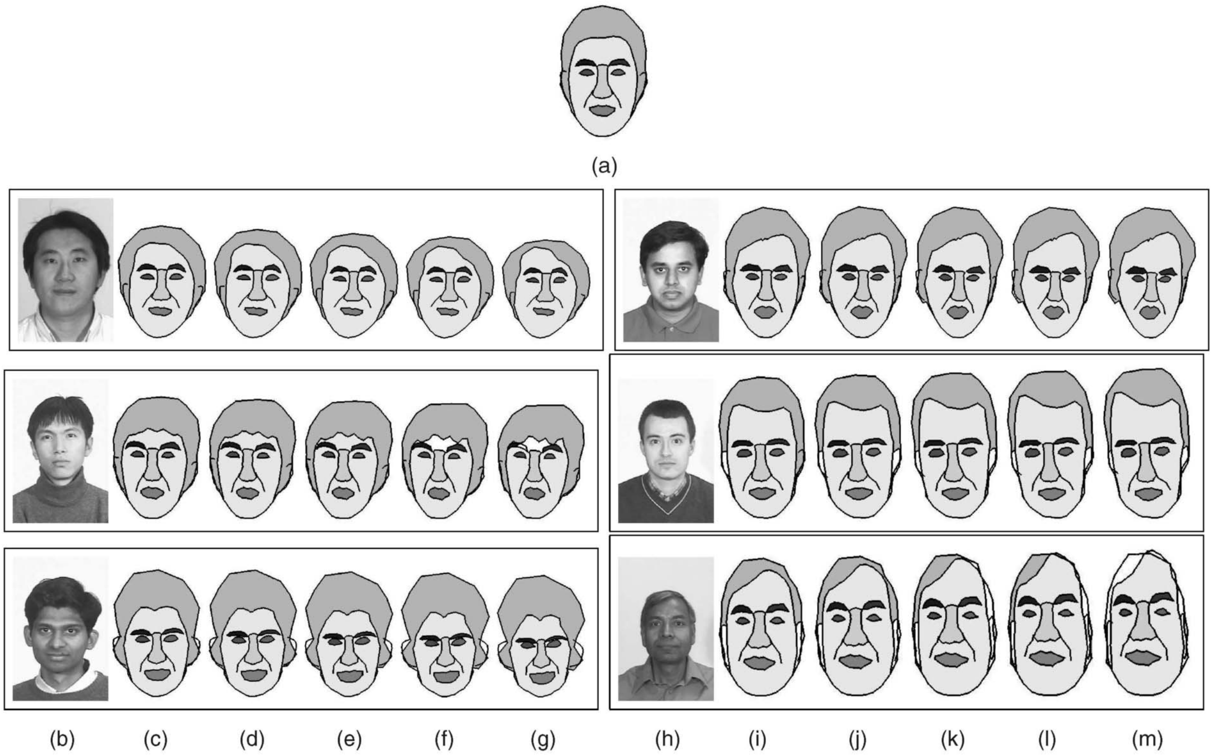

Fig. 10. Facial caricatures generated based on the average face of 50 faces (five for each subject): (a) A prototype of the semantic face graph, $\mathbf{G_0}$, obtained from the mean face of the database, with individual components shaded. (b) and (h) Face images of six different subjects. (c), (d), (e), (f), and (g) and (i), (j), (k), (l), and (m) Caricatures of the faces in (b) and (h), respectively, (semantic face graphs with individual components shown in different shades) with different values of exaggeration coefficients, k, ranging from 0.1 to 0.9.

$$SFD_i(\mathbf{G_0}, \mathbf{G_{P_k}}) = \mathbf{Dist}(\mathbf{SGD_i^{G_0}}, \mathbf{SGD_i^{G_{P_k}}}) = \left[\frac{\mathbf{1}}{\mathbf{L_i}}\sum_{\mathbf{k=0}}^{\mathbf{L_i}}\left|\mathbf{a_i^{G_0}(k)} - \mathbf{a_i^{G_{P_k}}(k)}\right|^{\mathbf{2}}\right]^{\mathbf{0.5}}, \quad (16)$$

where SGD stands for semantic graph descriptors. The distinctiveness of a facial component is evaluated by the semantic facial distance SFD between the generic semantic face graph and the aligned/matched semantic graph. The visibility of a facial component (due to head pose, illumination, and facial shadow) is estimated by the reliability of component matching/alignment (i.e., matching scores for facial components). Finally, the 2D semantic face graph of subject P can be learned from N_p images under similar pose by

$$\mathbf{G_P} = \bigcup_{\mathbf{i}} \mathcal{F}^{-1}\left\{\frac{\mathbf{1}}{\mathbf{N_P}}\sum_{\mathbf{k=1}}^{\mathbf{N_P}}\mathbf{SGD_i^{G_{P_k}}}\right\}. \quad (17)$$

```
INPUT:   - $N^j$ training face images for subject $P^j$, $j = 1, \ldots, M$
         - one query face image for unknown subject $Q$
Step 1:  Detect faces for all the images using the method in [9]
         $\longrightarrow$ Generate locations of face and facial features
Step 2:  Form a set of facial components, $T$, for recognition by assigning prior component weights
Step 3:  Coarsely align a generic semantic face graph to each image based on $T$
         $\longrightarrow$ Obtain component matching scores for each graph in Eq. (2)
Step 4:  Deform a coarsely-aligned face graph
         $\longrightarrow$ Update component matching scores based on integrals of component energies
Step 5:  Compute semantic facial descriptors $SGD$ for each graph
         using the 1-D Fourier transform in Eq. (1).
Step 6:  Compute semantic component weights for each graph in Eqs. (19)-(22)
Step 7:  Integrate all the face graphs of subject $P^j$ in Eq. (23),
         resulting in M template face graphs
Step 8:  Compute M matching costs, $C(P^j, Q_k)$, between $P^j$ and $Q_k$ in Eq. (24),
         where $k = 1$, $j = 1, \ldots, M$
Step 9:  Subject $P^J$ with the minimum matching cost has the best matched face to the unknown subject $Q_k$.
OUTPUT:  $Q = P^J$
```

Fig. 11. A semantic face matching algorithm.

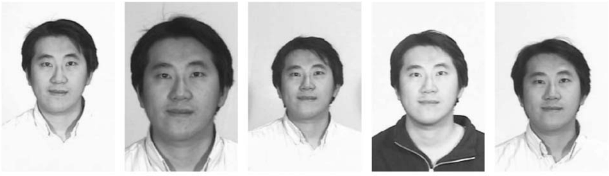

Fig. 12. Five color images (256×384) of a subject.

The matching cost between the subject P and the kth face image of subject Q can be calculated as

$$C(P, Q_k) = \sum_{i=1}^{M} \left\{ scw^{P}(i) \cdot scw^{Q_k}(i) \cdot SFD_i(\mathbf{G_P}, \mathbf{G_{Q_k}}) \right\}, \quad (18)$$

where M is the number of facial components. Face matching is accomplished by minimizing the matching cost.

5.2 Facial Caricatures

Facial caricatures are generated based on exaggeration of an individual's facial distinctiveness from the average facial topology. Let $\mathbf{G_P^{crc}}$ represent the face graph of a caricature for the subject P, and $\mathbf{G_0}$ be the face graph of the average facial topology. Caricatures are generated via the control of an exaggeration coefficient, k_i, in (19):

$$\mathbf{G_P^{crc}} = \bigcup_{\mathbf{i}} \mathcal{F}^{-1} \left\{ \mathbf{SGD_i}^{\mathbf{G_P}} + \mathbf{k_i} \cdot \left(\mathbf{SGD_i^{G_P}} - \mathbf{SGD_i^{G_0}} \right) \right\}. \quad (19)$$

Currently, we use the same coefficients for all the components, i.e., $k_i = k$. In Fig. 10, facial caricatures are optimized in the sense that the average facial topology is obtained from the mean facial topology of training images (a total of 50 images for 10 subjects). We can see that it is *easier* for a human to recognize a known face based on the exaggerated faces.

5.3 Face Matching

The proposed semantic face matching algorithm is described in Fig. 11 for face identification with no rejection. The computation of matching costs is based on the distance of semantic face descriptors and semantic component weights. We have constructed a small face database at near frontal views with small amounts of variations in facial expression, face orientation, face size, and lighting conditions, during different sessions over a period of two months. Fig. 12 shows five images of one subject, while Fig. 13 shows one image each of the 10 subjects. We employ five images each per subject for training and testing the semantic face graphs. With 5-fold cross validation tests, the cumulative rank score curves [17] are shown in Figs. 15a, 15b, 15c, and 15d using five different sets of facial components. External facial components include

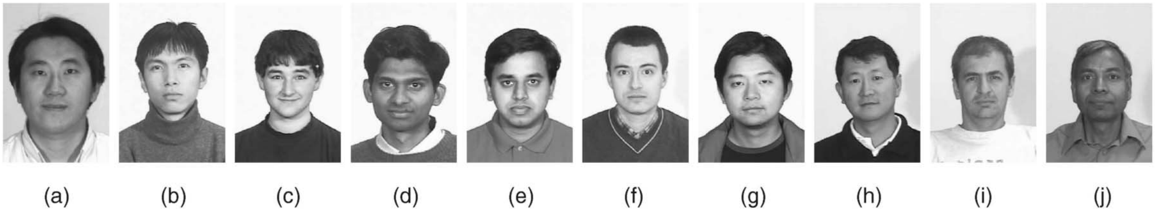

Fig. 13. Face images of 10 subjects.

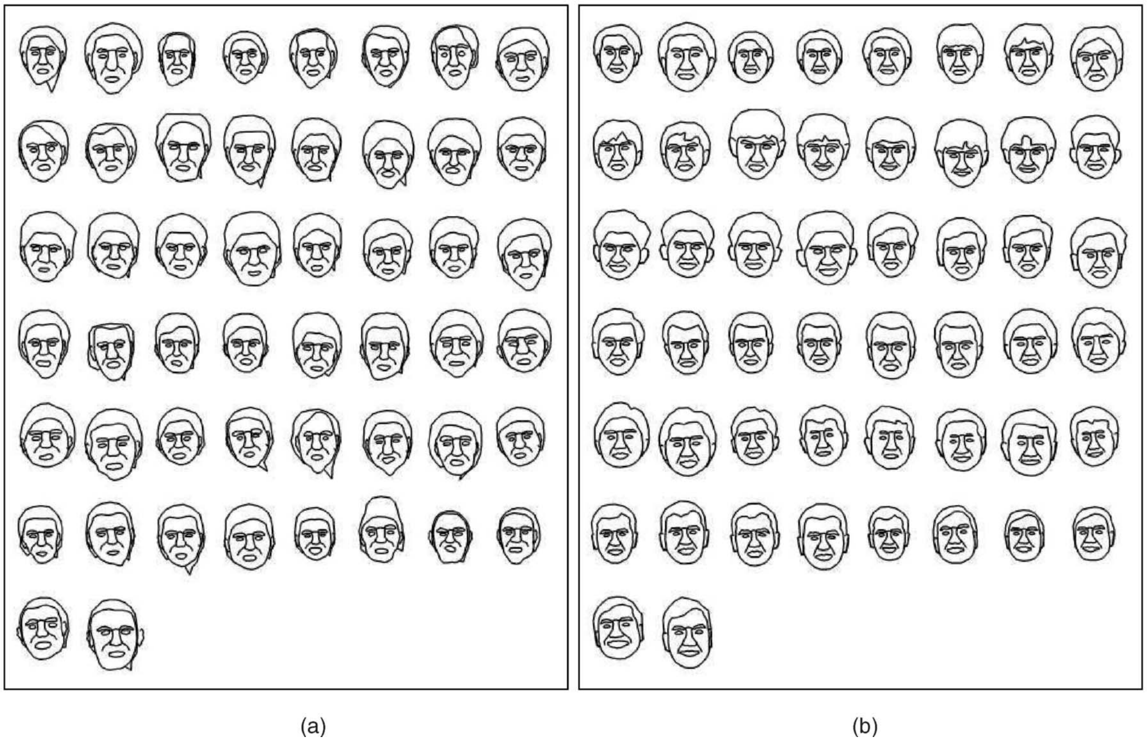

Fig. 14. Cartoon faces: (a) automatically aligned face graphs and (b) manually aligned graphs.

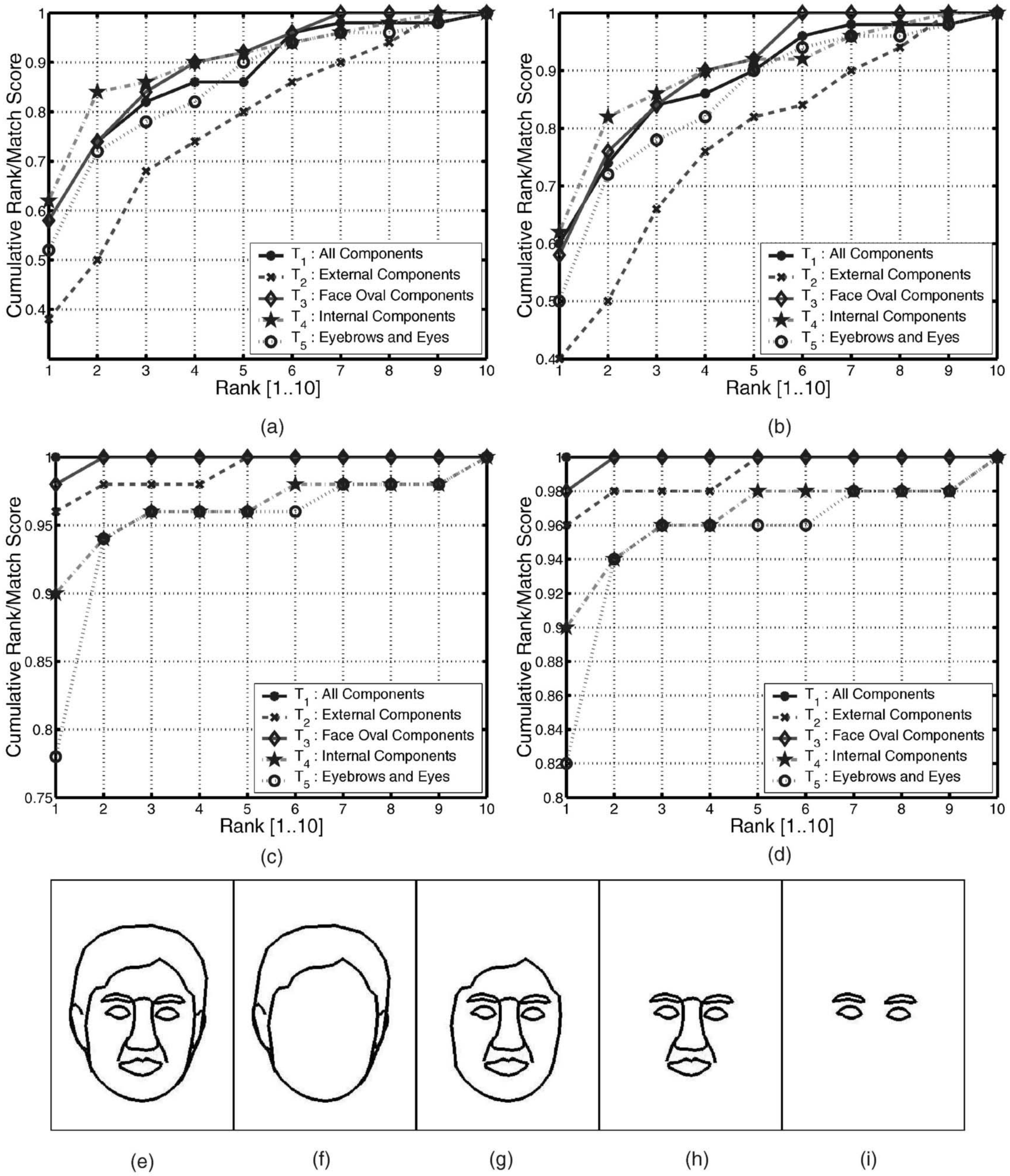

Fig. 15. Cumulative rank score curves obtained based on: (a) automatically aligned face graphs, (b) automatically aligned face graphs exaggerated with caricature scale $k = 0.7$, (c) manually aligned face graphs, and (d) manually aligned face graphs exaggerated with caricature scale $k = 0.7$. Five semantic sets are (e) all components (T_1), (f) external components (T_2), (g) face oval components (T_3), (h) internal components (T_4), and (i) eyes and eyebrows (T_5).

face outline, ears, and hairstyle, while internal components are eyebrows, eyes, nose, and mouth. With automatic face alignment (see Fig. 14a), the set of internal components gives the best performance (90 percent at the top four ranks). The set of external components are difficult to align accurately; therefore, it degrades the performance when all the components are used. However, with manual alignment (see Fig. 14b), we can see that the external facial components play an important role in recognition (resulting in a top-rank recognition rate of 96 percent and outperform the performance of internal components. The caricature exaggeration (see Figs. 15b and 15d) does improve the performance. The Fourier descriptors provide a compact feature set for classification and the dimensionality of the feature space is low (175 vertices for all the facial components). Each coarse alignment and fine alignment for an image of size 640×480 takes 6.84 secs (implemented in C) and 460 secs (implemented in MATLAB), repsectively, while each face comparison takes 0.0029 secs with Matlab implementation on a 1.7 GHz CPU. We are conducting other cross validation tests for classification, and are in the process of performing recognition on large gallery and probe databases. Although the alignment is currently done offline, we are attempting to improve both the alignment performance and alignment speed.

6 Conclusions and Future Work

We have proposed semantic face graphs derived from a subset of vertices of a 3D face model to construct cartoon faces for face matching. The cartoon faces are generated in a coarse-to-fine fashion; face detection results are used to coarsely align semantic face graphs with detected faces and interacting snakes are used to finely align face graphs with sensed face images. We have implemented an explicit snake deformation for fine alignment and shown that a successful interaction among multiple snakes associated with facial components makes the semantic face graph a useful model to represent faces. We have also presented a framework for semantic face recognition, which is designed to automatically derive weights for facial components based on their distinctiveness and visibility, and to perform face matching based on visible facial components. We have demonstrated good classification performance using extracted cartoon faces. An advantage of semantic face graph is that it allows face matching based on selected facial components and it also provides an effective way to update a 3D face model based on 2D images. We are currently adding snakes for ears and two open crest curves for the nose to complete the graph deformation of the entire face. In the future, we will evaluate the interacting snakes through two types of implementations, explicit (parametric active contours) and implicit (geodesic active contours) curve representations in the domain of face recognition. We plan to test the proposed semantic face matching algorithm on other face databases. We will also implement a pose estimation module in order to construct an automated pose-invariant face recognition system.

Acknowledgments

This research was supported by the Office of Naval Research grant no. N00014-01-1-0266 at Michigan State University.

References

[1] T. Abe and Y. Matsuzawa, "Multiple Active Contour Models with Application to Region Extraction," *Proc. 15th Int'l Conf. Pattern Recognition,* vol. 1, pp. 626-630, Sept. 2000.

[2] A.A. Amini, T.E. Weymouth, and R.C. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 9, pp. 855-867, Sept. 1990.

[3] S.E. Brennan, "Caricature Generator: The Dynamic Exaggeration of Faces by Computer," *Leonardo,* vol. 18, no. 3, pp. 170-178, 1985.

[4] V. Chalana, D.T. Linker, D.R. Haynor, and Y.M. Kim, "A Multiple Active Contour Model for Cardiac Boundary Detection on Echocardiographic Sequences," *IEEE Trans. Medical Imaging,* vol. 15, no. 3, pp. 290-298, 1996.

[5] T.F. Chan and L.A. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 266-277, 2001.

[6] C. Chesnaud, P. Réfrégier, and V. Boulet, "Statistical Region Snake-Based Segmentation Adapted to Different Physical Noise Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 11, pp. 1145-1157, Nov. 1999.

[7] I.J. Cox, J. Ghosn, and P.N. Yianilos, "Feature-Based Face Recognition Using Mixture-Distance," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 209-216, 1996.

[8] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky, "Fast Geodesic Active Contours," *IEEE Trans. Image Processing,* vol. 10, no. 10, pp. 1467-1475, 2001.

[9] R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face Detection in Color Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 696-706, May 2002.

[10] R.-L. Hsu and A.K. Jain, "Face Modeling for Recognition," *IEEE Int'l Conf. Image Processing,* vol. 2, pp. 693-696, Oct. 2001.

[11] J. Ivins and J. Porrill, "Statistical Snakes: Active Region Models," *Proc. Fifth British Machine Vision Conf.,* vol 2, pp. 377-386, 1994.

[12] W. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, no. 4, pp. 321-331, 1998.

[13] R. Mauro and M. Kubovy, "Caricature and Face Recognition," *Memory & Cognition,* vol. 20, no. 4, pp. 433-440, 1992.

[14] B. Olstad and A. H. Torp, "Encoding of A Priori Information in Active Contour Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 863-872, Sept. 1996.

[15] X.M. Pardo, M.J. Carreira, A. Mosquera, and D. Cabello, "A Snake for CT Image Segmentation Integrating Region and Edge Information," *Image and Vision Computing,* vol. 19, no. 7, pp. 461-475, 2001.

[16] P.S. Penev and J.J. Atick, "Local Feature Analysis: A General Statistical Theory for Object Representation," *Network: Computation in Neural Systems,* vol. 7, no. 3, pp. 477-500, 1996.

[17] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[18] G. Rhodes and T. Tremewan, "Understanding Face Recognition: Caricature Effects, Inversion, and the Homogeneity Problem," *Visual Cognition,* vol. 1, pp. 257-311, 1994.

[19] D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 6, pp. 569-579, June 1993.

[20] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[21] R.G. Uhl and N.d.V. Lobo, "A Framework for Recognizing a Facial Image from a Police Sketch," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 586-593, 1996.

[22] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.

[23] C.Y. Xu and J.L. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Trans. Image Processing,* vol. 7, no. 3, pp. 359-369, 1998.

[24] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah, "Shape from Shading: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 690-706, Aug. 1999.

[25] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," CVL Technical Report, Center for Automation Research, Univ. of Maryland at College Park, http://www.cfar.umd.edu/ftp/TRs/FaceSurvey.ps.gz, 2003.

[26] S.C. Zhu and A. Yuille, "Region Competition—Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.

Rein-Lien Hsu received the BSEE and MSEE degrees in electrical engineering from the National Cheng Kung University, Tainan, Taiwan, in 1990 and 1992, respectively, and the PhD degree in computer science and engineering from the Michigan State University in 2002. He is currently a senior research scientist at Identix Inc. , Jersey City, New Jersey. He specializes in the detection, modeling, and recognition of human faces, and 3D object reconstruction. His research interests include pattern recognition, signal and image processing, and computer vision. He is a member of the IEEE and the IEEE Computer Society.

Anil K. Jain is a University Distinguished Professor in the Department of Computer Science and Engineering at Michigan State University. He was the department chair between 1995 and 1999. His research interests include statistical pattern recognition, exploratory pattern analysis, Markov random fields, texture analysis, 3D object recognition, medical image analysis, document image analysis and biometric authentication. Several of his papers have been reprinted in edited volumes on image processing and pattern recognition. He received the best paper awards in 1987 and 1991, and received certificates for outstanding contributions in 1976, 1979, 1992, 1997, and 1998 from the Pattern Recognition Society. He also received the 1996 IEEE Transactions on Neural Networks Outstanding Paper Award. He is a fellow of the IEEE and International Association of Pattern Recognition (IAPR). He has received a Fulbright Research Award, a Guggenheim fellowship and the Alexander von Humboldt Research Award. He delivered the 2002 Pierre Devijver lecture sponsored by the International Association of Pattern Recognition (IAPR). He holds six patents in the area of fingerprint matching. His most recent book is *Handbook of Fingerprint Recognition*, Springer 2003.