SEMANTIC FACE MATCHING

Rein-Lien Hsu and Anil K. Jain

Dept. of Computer Science & Engineering, Michigan State University, MI 48824
Email: {hsureinl, jain}@cse.msu.edu

ABSTRACT

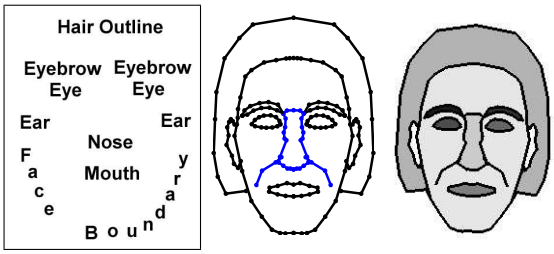
The need for efficient methods for archiving and retrieving personal digital photo collections arises due to a significant increase in the number of digital images and videos that people have to manage. We propose a semantic face matching approach for managing consumer photographs based on semantic face attributes. These attributes are organized as a semantic face graph (derived from a 3D generic face model) containing facial components such as eyes and mouth in the spatial domain. We align the semantic facial components in the semantic face graph with the extracted facial features in a given image. Aligned facial components are transformed to a feature space spanned by Fourier descriptors of facial components for face matching. The semantic face graph allows face matching based on selected facial components. Our experimental results demonstrate that the proposed semantic representation of the face is useful for face matching and visualization (e.g., generating facial caricatures).

1. INTRODUCTION

With the widespread use of digital cameras and camcorders and with the decrease in the cost of storage devices, consumers are facing the problem of managing huge collections of digital photos. Accompanying this increase in digital content is a need for database management tools that will allow people to easily archive and retrieve desired content from their digital collections. Furthermore, since humans and their activities are typically the subjects of interest in both images and videos [2], [3], [5], detection and identification of human faces will help to automate image and video archival based on semantic (high-level) concepts, such as the face and facial components. This will allow us to search a database using queries of the form "find all the images containing John's faces," and "search faces which have Vincent's eyes or Bill's chin." The traditional retrieval systems for digital visual content [4] based on text and low-level attributes (such as color, texture, shape, layout, and motion) can not process above-mentioned queries. Modeling facial components at a semantic level (i.e., eyebrows, eyes, nose, mouth, face outline, ears, and the hair outline) helps to reveal how the individual components contribute to face recognition. This will also allows us to capture the facial configuration, to determine distinctiveness of facial components, to assign local weights to facial components, and to separate external and internal facial components. We propose a *semantic* approach for face recognition which is based on 3D face model and semantic graph matching.

2. SEMANTIC FACE GRAPH

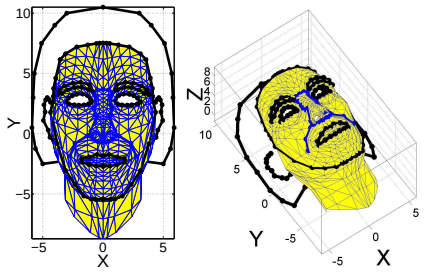
A semantic face graph provides a high-level description of the face and its facial components. A semantic graph in a frontal view is shown in Fig. 1. The nodes of the graph represent semantic facial components (e.g., eyes, mouth, and hair), each of which is constructed from a subset of connected vertices of the 3D generic face model. A semantic graph is represented in a 3D space and is compared with others in a 2D 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 [7] as the generic face



(a) (b) (c)

Figure 1: Semantic face graph shown in a frontal view and containing nodes (a) indicated by text; (b) depicted by curves; (c) filled with different shades. The edges of the semantic graph are implicitly stored in a 3D generic face model and are hidden here.

model because it contains all the internal facial components, and the face outline, and muscle models for mimicking facial expressions. However, Waters' mesh model does not include the external facial features such as ears and hair. The hair and the face outline play a crucial role in face recognition. Hence, we add these external facial components to Waters' model (currently, only for the frontal view). We decompose the vertices of the mesh model into two sets:



(a) (b)

Figure 2: 3D triangular-mesh model overlaid with facial curves including hair and ears at a (a) frontal view; (b) side view.

(i) vertices at the boundaries of facial components and (ii) vertices belonging to facial component regions. The vertices of a component boundary are described by polygonal curves (see Fig. 2). Let $\mathbf{T_0}$ denote the set of all semantic facial components, which are nodes of the generic semantic graph, $\mathbf{G_0}$. That is T_0 ={{left eyebrow}}, {right eyebrows}, {left eye}, ..., {hair boundary} }. Let T be a subset of T_0, that is $T \subset 2^{T_0}$. Let M be the number of facial components. For example, T can be specified as {{left eye}, {right eye}, {mouth}}, where M is 3. Let the semantic graph projected on a 2D image represented by the set $\mathbf{T}$ be $\mathbf{G}$. The boundary coordinates 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 [9], $a_i(k)$, of the signal $u_i(n) = x_i(n) + jy_i(n)$ is computed as

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

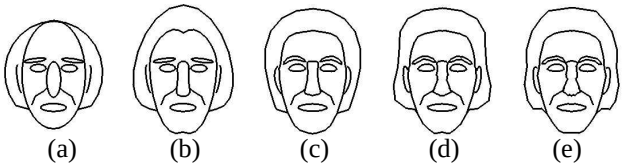
for facial component i with a close boundary such as eyes and mouth, and with end-vertex padding for those having open boundary such as ears and hair components. The advantage of using semantic graph descriptors for face matching is that these descriptors can seamlessly encode geometric relationships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

$$\tilde{u}_i(n) = \mathcal{F}^{-1}\{a_i(k)\} = \sum_{k=0}^{L_i-1} a_i(k) \cdot e^{j2\pi kn/N_i}, \quad (2)$$

where $L_i (< N_i)$ is the number of frequency components used for component i. Figure 3 shows the reconstructed semantic face graphs at different levels of Fourier series truncation.

3. FACE RETRIEVAL SYSTEM

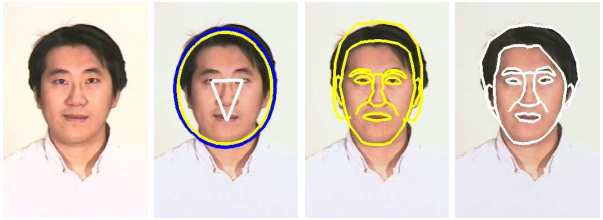
We propose a face retrieval system that contains four major modules: face detection, pose estimation, face alignment,



(a) (b) (c) (d) (e)

Figure 3: Semantic face graphs at the frontal view are reconstructed using Fourier descriptors with spatial frequency components increasing from (a) 10% to (e) 100%.

and face matching. The face detection module locates faces (in non-profile views) and facial components in a color image using the algorithm in [1]. Figures 4 (a) and (b) show an input color image and the detection results. Currently, we assume that the face images have been captured at near frontal views; we are implementing a pose estimation module. The alignment module uses the face detection results to align a semantic face graph and the input image in a coarse-to-fine fashion. In the coarse alignment, a semantic face



(a) (b) (c) (d)

Figure 4: Face alignment: (a) input color image; (b) detected face that is described by an ellipse and an eye-mouth triangle; (c) and (d) coarsely and finely aligned semantic face graphs, respectively, overlaid on the face image.

graph is aligned with a given face through the *global* scaling, rotation, and translation, based on the detected locations of the face and facial components. In the fine alignment, the semantic face graph is *locally* deformed to fit the face using multiple snakes. Figures 4(c) and 4(d) show the results after the coarse and fine alignment, repsectively. In this paper, the fine alignment is based on the use of a semi-automatic graphical user interface, although we have automated the fine alignment for the hair and face outlines, eyes, nose, and mouth. After aligning semantic face graph to an image, a matching score for each facial component is generated according to the visibility of individual components.

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

$$scw^P(i) = \begin{cases} 1 + e^{-2\sigma_d^2(i)/d^2(i)} & N_p > 1, \\ 1 + e^{-1/d^2(i)} & N_p = 1, \end{cases} \quad (3)$$

$$d(i) = \frac{1}{N_P} \sum_{k=1}^{N_P} SFD_i(\mathbf{G_0}, \mathbf{G_{P_k}}) \cdot ms^{P_k}(i), \quad (4)$$

$$\sigma_d(i) = SD_k\left[SFD_i(\mathbf{G_0}, \mathbf{G_{P_k}}) \cdot ms^{P_k}(i)\right], \quad (5)$$

where SFD means semantic facial distance, ms is matching score, SD stands for standard deviation, $\mathbf{G_0}$ 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 between two graphs for facial component i is defined as

$$SFD_i(\mathbf{G_0},\mathbf{G_{P_k}}) = Dist(SGD_i^{\mathbf{G_0}}, SGD_i^{\mathbf{G_{P_k}}}) = \left[\frac{1}{L_i}\sum_{k=0}^{L_i}\left|a_i^{\mathbf{G_0}}(k) - a_i^{\mathbf{G_{P_k}}}(k)\right|^2\right]^{0.5}, \quad (6)$$

where SGD stands for semantic graph descriptor. 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 matching reliability (i.e., matching scores for facial components). Finally, the 2D semantic face graph of subject P can be learned from N_p images captured under the similar pose by

$$\mathbf{G_P} = \bigcup_i \mathcal{F}^{-1}\left\{\frac{1}{N_P}\sum_{k=1}^{N_P} SGD_i^{\mathbf{G_{P_k}}}\right\}. \quad (7)$$

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

$$C(P,Q_k) = \sum_{i=1}^{M}\left\{scw^P(i)\cdot scw^{Q_k}(i)\cdot SFD_i(\mathbf{G_P},\mathbf{G_{Q_k}})\right\}, \quad (8)$$

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

4. FACE MATCHING

We have constructed our color 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. Figure 5 shows five images of one subject, while Fig. 6 shows one image of each of ten subjects. We employ 5 images per subject for training the semantic face graphs. With re-substitution and leave-one-out tests, the misclassification rates are shown in Table 1 using different sets of facial components and semantic graph descriptors with the number of frequency components truncated at three different levels. Classification errors might be due to the lack of local texture information in matching. External facial components are ears, and the hair and the face outlines, while internal components are eyebrows, eyes, nose, and mouth. Table 1 shows that the external facial components play an important role in recognition, and the Fourier descriptors provide compact features for classification because the dimensionality of the feature space is lower (see Table 2), compared to that used in eigen-subspace methods. Figures 7 and 8 show the reconstructed semantic face graphs, $\mathbf{G_P}$ in Eq. (7), (compare them with $\mathbf{G_0}$ in Fig. 1(c)) at two levels of detail. Each face comparison takes 0.0029 sec with Matlab implementation on a 1.7 GHz CPU. We are conducting other cross-validation tests for the classification, and are performing recognition on gallery and probe databases. The semantic face graph is useful for generating caricatures based on the component distinctiveness (see Fig. 9).

Figure 5: Five color images (256×384) of one subject.

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 6: One face image of each of ten subjects.

Table 1: Error rates on a 50-image database.

Set	T_1		T_2		T_3		T_4	
Graph								
P (%)	RS	LOO	RS	LOO	RS	LOO	RS	LOO
100%	0%	6%	0%	6%	12%	24%	16%	30%
50%	0%	6%	0%	6%	12%	24%	16%	30%
30%	0%	6%	0%	12%	16%	24%	18%	34%

P: % of frequency components, T_1: All components, T_2: External components, T_3: Internal components, T_4: Eyes and Eyebrows, RS: Re-substitution, LOO: Leave-one-out.

5. CONCLUSIONS AND FUTURE WORK

We have presented the framework of 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 these component

weights. The proposed semantic face graph is useful for (i) constructing compact features for face classification that can incorporate geometric face shape through Fourier descriptors, (ii) face matching based on selected components, and (iii) generating facial caricatures. We plan to fully automate the deformation of semantic graph for different head poses, incorporate component texture (shadings) into the semantic graph descriptors, update the 3D face model based on 2D images, and enlarge the color face database for classification.

Table 2: Dimensionality of semantic graph descriptors for individual facial components.

P (%)	100%	50%	30%
Dimension	N_i	L_i	L_i
Eyebrow	12	5	3
Eye	13	7	3
Nose	34	13	7
mouth	14	7	3
Face outline	36	17	11
Ear	11	5	3
Hair	19	9	5

P: % of frequency components, N_i: the dimension of semantic graph descriptors, L_i: the dimension of truncated descriptors.

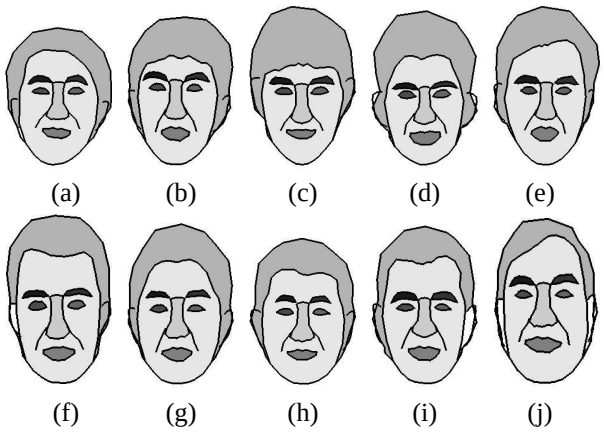


Figure 7: Semantic face shapes of 10 subjects reconstructed from Fourier descriptors using all the frequency components.

6. REFERENCES

[1] 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.

[2] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face database," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 132–137, 2000.

[3] A. Martinez, "Face Image Retrieval Using HMMs," *Proc. IEEE Workshop Content-Based Access of Image and Video Libraries*, pp. 25–39, June 1999.

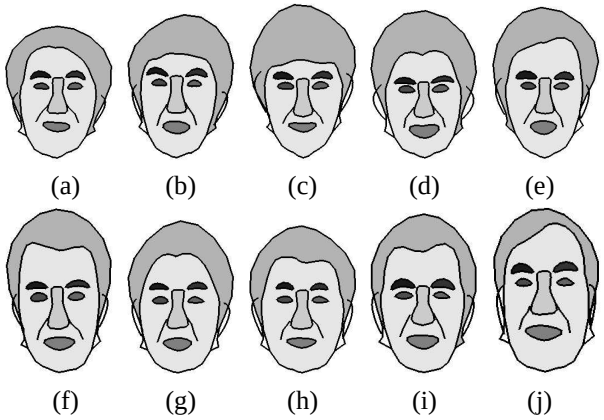


Figure 8: Semantic face shape reconstructed from Fourier descriptors using only 50% of frequency components.

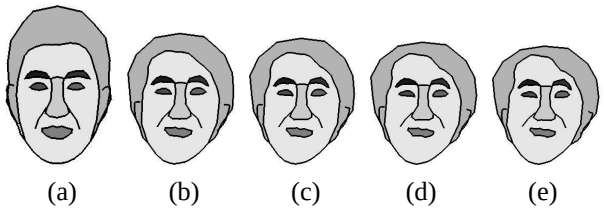


Figure 9: Facial caricatures of one subject: (a) average face graph of the 50 faces (5 for each subject), with individual components shaded; (b)–(e) caricatures of the face in Fig. 6(a) (i.e., semantic face graphs shown with increasing distortion between the graph in 9(a) and the graph in Fig. 7(a).

[4] A.W.M. Smeulders, M. Worring, S. Santini, and A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[5] L. Torres and J. Vila, "Automatic face recognition for video index applications," *Pattern Recognition*, vol. 35, no. 3, pp. 615–625, Mar. 2002.

[6] J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, Sep. 2001.

[7] F.I. Parke and K. Waters, "Appendix 1: Three-dimensional muscle model facial animation," *Computer Facial Animation*, A.K. Peters, 1996.

[8] H. Wechsler, J. P. Phillips, V. Bruce, F. Fogelman-Soulie and T. Huang (Eds.), *Face Recognition: From Theory to Applications*, Springer-Verlag, 1998.

[9] C.T. Zahn and R.Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Computers*, vol. C-21, no. 3, pp. 269–281, 1972.

[10] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," *CVL Technical Report TR4167*, Center for Automation Research, University of Maryland at College Park, Oct. 2000.