MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Bayesian Face Recognition with Deformable Image Models

Baback Moghaddam [*]
Chahab Nastar [†]
Alex Pentland [‡]

TR2001-05   September 2001

Abstract

We propose a novel representation for characterizing image differences using a deformable technique for obtaining pixel-wise correspondences. This representation, which is based on a deformable 3D mesh in XYI-space, is then experimentally compared with two related correspondence methods: optical flow and intensity differences. Furthermore, we make use of a probabilistic similarity measure for direct image matching based on a Bayesian analysis of image variations. We model two classes of variation in facial appearance: *intra-personal* and *extra-personal*. The probability density function for each class is estimated from training data and used to compute a similarity measure based on the *a posteriori* probabilities. The performance advantage of our deformable probabilistic matching technique is demonstrated using 1700 faces from the US Army's "FERET" face database.

International Conference on Image Analysis & Processing (ICIAP'01),
Palermo, Italy, September 2001

[*] MERL - Research Laboratory
[†] INRIA Rocquencourt
[‡] MIT Media Laboratory

Published in: *International Conference on Image Analysis & Processing*, (ICIAP'01), 2001.

Bayesian Face Recognition with Deformable Image Models

Baback Moghaddam[1], Chahab Nastar[2], and Alex Pentland[3]

[1] Mitsubishi Electric Research Laboratory, Cambridge, MA 02139, USA.
[2] INRIA Rocquencourt, B.P. 105, F-78153 Le Chesnay Cédex, France.
[3] MIT Media Laboratory, 20 Ames St., Cambridge, MA 02139, USA.

Abstract

We propose a representation for characterizing facial image differences using a deformable technique for obtaining pixel-wise correspondences. This representation, which is based on a deformable 3D mesh in XYI-space, is then experimentally compared with two related correspondence methods: optical flow and intensity differences. Furthermore, we use a probabilistic similarity measure for matching based on a Bayesian analysis of image variations. We model two classes of variation in facial appearance: intra-personal *and* extra-personal. *The probability density function for each class is estimated from training data and used to compute a similarity measure based on the* a posteriori *probabilities. The performance advantage of our deformable probabilistic matching technique is demonstrated using 1700 faces from the US Army's "FERET" face database.*

1 Introduction

In its simplest form, the similarity measure $S(I_1, I_2)$ between two images I_1 and I_2 can be set to be inversely proportional to the norm $||I_2 - I_1||$. Such a simple formulation suffers from two major drawbacks: it requires precise alignment of the objects in the image and does not exploit knowledge of which type of variations are critical (as opposed to incidental) in expressing similarity. In this paper, we use a *probabilistic* similarity measure based on the probability that the image-based differences, denoted by $d(I_1, I_2)$, are characteristic of typical variations in appearance of the *same* object. For example, for purposes of face recognition, we can define two classes of facial image variations: *intrapersonal* variations Ω_I (corresponding, for example, to different facial expressions of the *same* individual) and *extrapersonal* variations Ω_E (corresponding to variations between *different* individuals). Our similarity measure is then expressed in terms of the probability

$$S(I_1, I_2) \;=\; P(\Omega_I \mid d(I_1, I_2)) \qquad (1)$$

where $P(\Omega_I \mid d(I_1, I_2))$ is the *a posteriori* probability given by Bayes rule, using estimates of the likelihoods $P(d(I_1, I_2) \mid \Omega_I)$ and $P(d(I_1, I_2) \mid \Omega_E)$ which are derived from training data using an efficient subspace method for density estimation of high-dimensional data [9]. In addition to the use of this probabilistic simiarlity measure, we explore a novel representation for $d(I_1, I_2)$ which corresponds to the *parametric modes* of a deformable intensity surface. We believe that this representation affords a convenient and unifying mathematical framework for incorporating both the 2D *shape* and *texture* components of an object for visual recognition. We propose a novel representation for $d(I_1, I_2)$ which combines both the spatial (XY) and grayscale (I) components of the image in a unified XYI framework consisting of a deformable surface mesh.

2 Deformable Surfaces

The idea of using intensity surfaces for matching and recognition comes from the observation that the transformation of shape to intensity is quasi-linear under controlled lighting conditions ; in other words, the intensity of the 2D image reflects the actual 3D shape. This essential observation is the basis of all shape from shading methods [5] ; however, unlike those methods, our aim is not actually to reconstruct depth information from a single 2D projection, but rather note that under controlled lighting conditions the changes in the image intensities from one image to the other reflect changes in their actual 3D shape [13]. Mathematically, assuming the object of interest to be a Lambertian (or matte) surface, the amount of intensity reflected when illuminated by a single light

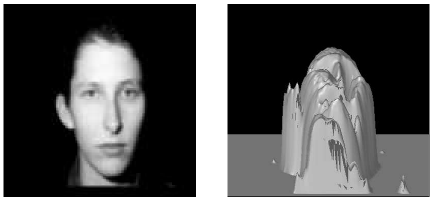


Figure 1: An image and its XYI surface representation

source placed at infinity, is isotropic :

$$I(x,y) = \alpha\,\vec{N}(x,y).\vec{L} \tag{2}$$

where $\vec{N}(x,y)$ is the surface normal vector at point (x,y), $\vec{L}$ is the light source vector, and α is a positive scalar. This equation directly links shape $\vec{N}(x,y)$ and intensity surface $I(x,y)$ (figure 1). If the shape is relatively smooth, we can represent the image intensity as a continuous surface:

$$(x,y) \longrightarrow I(x,y) \tag{3}$$

This paper focuses on statistical analysis for recognition in the 3D space defined by $(x, y, I(x,y))$, which we will call the XYI space.

2.1 XYI Warping

Following the theory of active contour models [7, 18], several models have been developed that deal explicitly with deformable surfaces, among them : deformable superquadrics [14, 17], surface snakes [3, 8], particle systems [16], splines [2] and elastic thin plates [15, 12]. The above models usually evolve in Euclidean 3D space, however, deformable templates which evolve in XYI space with application to feature extraction have been investigated by Yuille *et al* [20]. Hence, deformable intensity surfaces with application to face recognition is a new approach to matching and recognition.

The mathematical approach to our model is inspired by the one described in [12]. The intensity surface is modeled as a deformable mesh of N nodes and is governed by Lagrangian dynamics [1] :

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}(t) \tag{4}$$

where $\mathbf{U} = [\ldots, \Delta x_i, \Delta y_i, \Delta z_i, \ldots]^T$ is a vector storing nodal displacements, $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are respectively the mass, damping and stiffness matrices of the system, and $\mathbf{F}$ is the external force. The above equation is of order $3N$ coresponding to the three displacement directions X, Y, I.

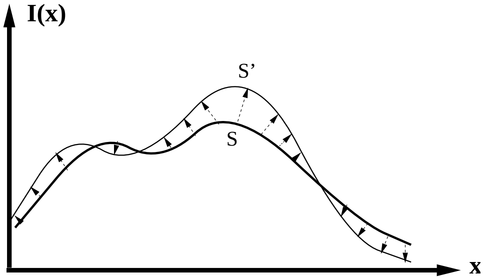


Figure 2: A cross-section of the intensity surface S being pulled towards S' by image forces

In warping one image onto a second (reference) image, the external force at each node P_i of the mesh is the vector to the closest 3D point Q_i in the reference surface:

$$\mathbf{F}(t) = [\ldots, \overrightarrow{P_iQ_i}(t), \ldots]^T \tag{5}$$

Euclidean distance algorithms can help us extract this force in each voxel of the 3D image, as a pre-processing [4, 19]. The final correspondence (and consequently the resultant XYI-warp) between two images is obtained by solving the governing equation above. Figure 2 shows a schematic representation of the deformation process. Note that the external forces (dashed arrows) do *not* necessarily correspond to the final displacement field of the surface. The elasticity of the surface provides an intrinsic smoothness constraint for computing the final displacement field.

2.2 Modal Analysis

Equation 4 is an impractically large matrix equation to solve. Instead modal analysis seeks to jointly diagonalize the mass and stiffness matrices in the new (modal) coordinate system. The vibration modes $\boldsymbol{\phi}(i)$ of the deformable surface are then the vector solutions of the eigenproblem:

$$\mathbf{K}\boldsymbol{\phi} = \omega^2\mathbf{M}\boldsymbol{\phi} \tag{6}$$

where $\omega(i)$ is the i-th eigenfrequency of the system. This eigen-decomposition yields, in modal coordinates, $\tilde{\mathbf{K}} = diag(\ldots\omega_i\ldots)$ and $\tilde{\mathbf{M}} = \mathbf{I}$. Consequently, $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$ is also diagonalized to $\tilde{\mathbf{C}} = diag(\ldots\tilde{c}_i\ldots)$. Solving the governing equations in the modal basis then leads to scalar equations where the unknown $\tilde{u}(i)$ is the amplitude of the deformation mode i [1]

$$\ddot{\tilde{u}}(i) + \tilde{c}_i\dot{\tilde{u}}(i) + \omega(i)^2\tilde{u}(i) = \tilde{f}_i(t) \quad i = 1, \ldots, 3N \tag{7}$$

In particular, for surface meshes, each mode is defined by two parameters $(i = (p, p'))$. The closed-form

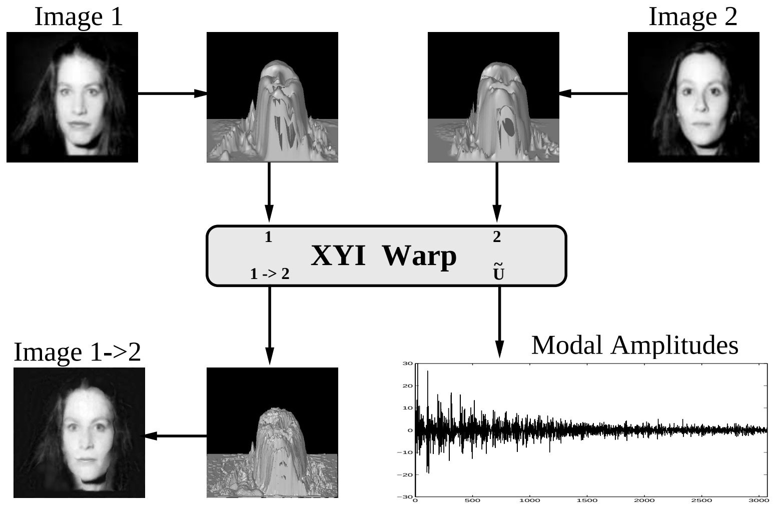


Figure 3: Example of XYI warping two images.

expression of the displacement field is then given by

$$\mathbf{U} \approx \sum_{i=1}^{P} \tilde{u}(i)\boldsymbol{\phi}(i) \tag{8}$$

with $P \ll 3N$, which means that only P scalar equations of the type (7) need to be solved. The modal superposition equation (8) can be seen as a Fourier expansion with high-frequencies neglected [11]. In our formulation, however, we make use of the *analytic modes* [11, 13], which are known sine and cosine functions for specific surface topologies

$$\boldsymbol{\phi}(p,p') = [\ldots, \cos\frac{p\pi(2i-1)}{2n}\cos\frac{p'\pi(2j-1)}{2n'}, \ldots]^T \tag{9}$$

These analytic expressions avoid costly eigenvector decompositions and furthermore allow the total number of modes to be easily adjusted for the application. We note that the above modal analysis technique represents a coordinate transform from the nodal displacement space to the modal amplitude subspace:

$$\tilde{\mathbf{U}} = \mathbf{\Phi}^T\mathbf{U} \tag{10}$$

where $\mathbf{\Phi}$ is the matrix of analytic modes $\boldsymbol{\phi}(p,p')$ and $\tilde{\mathbf{U}}$ is the resultant vector of modal amplitudes which encodes the type of deformations which characterize the difference between the two images.

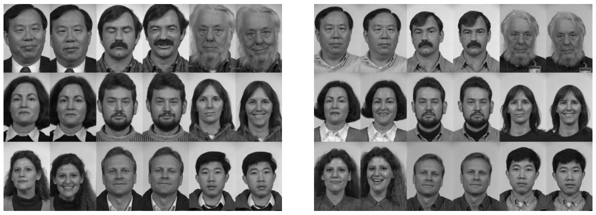
3 Analysis of Deformations

Consider the problem of characterizing the type of deformations which occur when matching two images in a face recognition task. We define two distinct and mutually exclusive classes: Ω_I representing *intrapersonal* variations between multiple images of the same individual (*e.g.,* with different expressions and lighting conditions), and Ω_E representing *extrapersonal* variations which result when matching two different individuals. We will assume that both classes are Gaussian-distributed and seek to obtain estimates of the likelihood functions $P(\tilde{\mathbf{U}}|\Omega_I)$ and $P(\tilde{\mathbf{U}}|\Omega_E)$ for a given deformation's modal amplitude vector $\tilde{\mathbf{U}}$. Given these likelihoods we can define the similarity score $S(I_1, I_2)$ between a pair of images directly in terms of the intrapersonal *a posteriori* probability as given by Bayes rule:

$$S(I_1, I_2) = \frac{P(\tilde{\mathbf{U}}|\Omega_I)P(\Omega_I)}{P(\tilde{\mathbf{U}}|\Omega_I)P(\Omega_I) + P(\tilde{\mathbf{U}}|\Omega_E)P(\Omega_E)} \tag{11}$$

3.1 Statistical Modeling of Modes

One difficulty with this approach is that the modal amplitude vectors are high-dimensional — $\tilde{\mathbf{U}} \in \mathcal{R}^N$ with $N = O(10^3)$. Therefore we typically lack sufficient independent training observations to compute reliable 2nd-order statistics for the likelihood densities (*i.e.,* singular covariance matrices will result). An efficient density estimation method for such a case was proposed by Moghaddam & Pentland [10] which divides the vector space $\mathcal{R}^N$ into two complementary subspaces using an eigenspace decomposition. This method relies on Principal Components Analysis (PCA) [6] to form a low-dimensional estimate of the complete likelihood which can be evaluated using only the first M principal components, where $M << N$. This decomposition



(a) (b)

Figure 4: Examples of FERET frontal-view image pairs used for (a) the Gallery set (training) and (b) the Probe set (testing).

forms an orthogonal decomposition of the vector space $\mathcal{R}^N$ into two mutually exclusive subspaces: the principal subspace F containing the first M principal components and its orthogonal complement $\bar{F}$, which contains the residual of the expansion. As shown in [10], the complete likelihood estimate can be written as the product of two independent marginal Gaussian densities

$$\begin{aligned}\hat{P}(\tilde{\mathbf{U}}|\Omega) &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\cdot\left[\frac{\exp\left(-\frac{\epsilon^2(\tilde{\mathbf{U}})}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \\ &= P_F(\tilde{\mathbf{U}}|\Omega)\;\hat{P}_{\bar{F}}(\tilde{\mathbf{U}}|\Omega)\end{aligned} \tag{12}$$

where $P_F(\tilde{\mathbf{U}}|\Omega)$ is the true marginal density in F, $\hat{P}_{\bar{F}}(\tilde{\mathbf{U}}|\Omega)$ is the estimated marginal density in the orthogonal complement $\bar{F}$, y_i are the principal components and $\epsilon^2(\tilde{\mathbf{U}})$ is the residual (or DFFS). The optimal value for the weighting parameter ρ is simply the average of the $\bar{F}$ eigenvalues

$$\rho = \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i \tag{13}$$

4 Experiments

Our experimental data consisted of a training "gallery" of 700 individual FERET faces and 1000 test images or "probes" containing one or more images of every person in the gallery. This collection of images consists of typically hard recognition cases that have proven difficult mainly due to the fact that the images were taken at different times, at different locations, and under different imaging conditions. Representative (unprocessed) images are shown in Figure 4. Before applying our deformable matching technique, we performed a rigid alignment of the facial images using an automatic face-processing system which extracts faces from the input image and normalizes for translation, scale as well as slight rotations. This system is described in detail in Moghaddam & Pentland [10] and uses maximum-likelihood estimation of object location — see Figure 5.

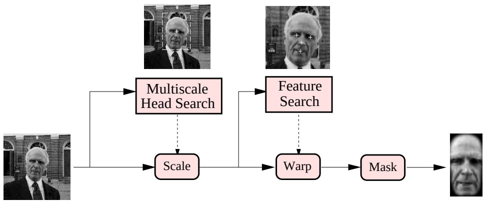


Figure 5: The face alignment system

4.1 Matching with Eigenfaces

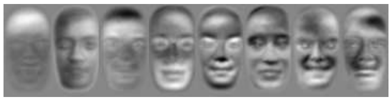
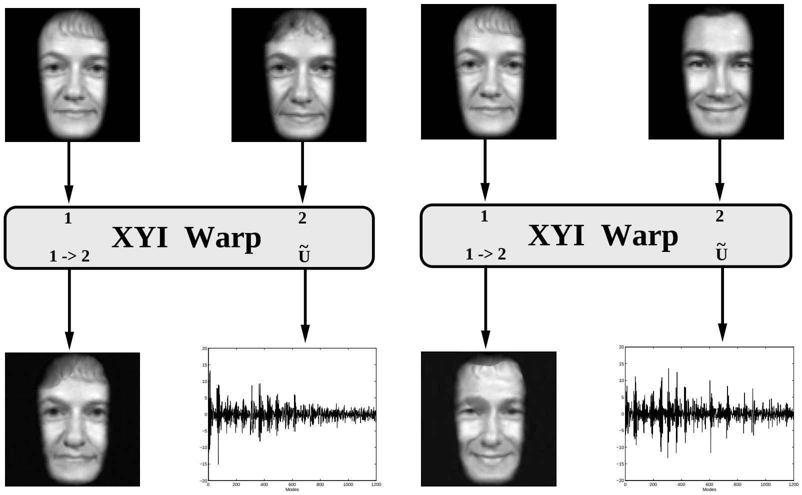


Figure 6: The first 8 normalized eigenfaces.

As a baseline comparison, we first used an eigenface matching technique for recognition. The normalized images from the gallery and the probe sets were projected onto a 100-dimensional eigenspace and a nearest-neighbor rule based on a Euclidean distance measure was used to match each probe image to a gallery image. A few of the lower-order eigenfaces used for this projection are shown in Figure 6. We note that these eigenfaces represent the principal components of an entirely different set of images — *i.e.,* none of the individuals in the gallery or probe sets were used in obtaining these eigenvectors. In other words, neither the gallery nor the probe sets were part of the "training set." The rank-1 recognition rate obtained with this method was found to be 79.5% and the correct match was always in the top 10 nearest neighbors.

4.2 Matching with XYI Warps

For our probabilistic algorithm, we first gathered training data by computing the modal amplitude spectra for a training subset of 700 intrapersonal warps (by matching the two views of every individual in the gallery) and a random subset of 1500 extrapersonal warps (by matching images of *different* individuals in the gallery), corresponding to the classes Ω_I and Ω_E, respectively. An example of each of these two



(a) (b)

Figure 7: Examples of (a) intrapersonal and (b) extrapersonal facial warps.

types of warps is shown in Figure 7. We computed the likelihood estimates $P(\tilde{\mathbf{U}}|\Omega_I)$ and $P(\tilde{\mathbf{U}}|\Omega_E)$ using the PCA-based method outlined in Section 3.1. We selected principal subspace dimensions of $M_I = 10$ and $M_E = 30$ for Ω_I and Ω_E, respectively. These density estimates were then used with a default setting of equal priors, $P(\Omega_I) = P(\Omega_E)$, to evaluate the *a posteriori* intrapersonal probability $P(\Omega_I|\tilde{\mathbf{U}})$ for matching the 1000 probe images to the 700 in the gallery. Therefore, for each probe image we computed all probe-to-gallery warps and sorted the matching order using the *a posteriori* probability $P(\Omega_I|\tilde{\mathbf{U}})$ as the similarity measure. This probabilistic ranking yielded a peak rank-1 recognition rate of 97.8%.

4.3 Matching with Optical Flow and Intensity Differences

To compare our deformable representation for $d(I_1, I_2)$ (*i.e.,* the modal amplitudes of an XYI-warp), we next applied our Bayesian matching technique on the alternative representations: intensity differences and optical flow. For each method, the eigenspace analysis was used to derive corresponding density estimates for the intra/extra classes and recognition proceeded exactly as described in the previous section. Since it is difficult to compare recognition and false match rates directly (due to the different dimensionalities of $d(I_1, I_2)$ in each case) we systematically varied the dimensions of the principal subspaces M_I and M_E, for each method and analyzed the performance in terms of % correct recognition. Table 8 shows the results averaged over nearly 2,000 different combinations of M_I and M_E for the three different methods: full XYI-warp, intensity differences (I-diff) and optical flow (XY-flow). These results indicate that XYI-warps are in fact the best representation for classification purposes, with intensity differences being second and optical flow being the least effective representation.

5 Conclusions

We have argued in favor of a *probabilistic* measure of facial similarity, in contrast to simpler methods which are based on standard L_2 norms (*e.g.,* template matching) or subspace-restricted norms (*e.g.,* eigenspace matching). This probabilistic framework is also advantageous in that the intra/extra density estimates explicitly characterize the type of appearance variations which are critical in formulating a meaningful measure of similarity. Furthermore, we have experimentally shown that our deformable XYI warping method for obtaining pixel correspondences does indeed lead to an effective representation especially when compared with simpler methods such as intensity differences and optical flow.

References

[1] K. J. Bathe. *Finite Element Procedures in Engineering Analysis.* Prentice-Hall, 1982.

[2] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(6):567–585, June 1989.

	XYI-warp	**I-diff**	**XY-flow**
Recognition Rate	**96.3 ± 2.51%**	**90.5 ± 2.34 %**	**83.8 ± 4.57%**

Figure 8: Performance of Bayesian classifier with three different data representations: full XYI-warp, intensity differences (I-diff) and optical flow (XY-flow). Based on nearly 2000 experimental trials with varying M_I and M_E.

[3] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-15(7):1131–1147, 1993.

[4] P. E. Danielsson. Euclidean distance mapping. *Computer Vision, Graphics, and Image Processing*, 14:227–248, 1980.

[5] B.K.P. Horn. *Robot Vision.* McGraw-Hill, New York, 1986.

[6] I. T. Jolliffe. *Principal Component Analysis.* Springer-Verlag, New York, 1986.

[7] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[8] T. McInerney and D. Terzopoulos. A finite element model for 3-D shape reconstruction and nonrigid motion tracking. In *IEEE Proceedings of the Fourth International Conference on Computer Vision (ICCV'93)*, pages 518–523, Berlin, June 1993. IEEE.

[9] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE Proceedings of the Fifth International Conference on Computer Vision (ICCV'95)*, Cambridge, USA, June 1995.

[10] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-19(7):696–710, July 1997.

[11] C. Nastar. Vibration modes for nonrigid motion analysis in 3D images. In *Proceedings of the Third European Conference on Computer Vision (ECCV '94)*, Stockholm, May 1994.

[12] C. Nastar and N. Ayache. Fast segmentation, tracking, and analysis of deformable objects. In *IEEE Proceedings of the Third International Conference on Computer Vision (ICCV'93)*, Berlin, May 1993.

[13] C. Nastar and A. Pentland. Matching and recognition using deformable intensity surfaces. In *IEEE International Symposium on Computer Vision*, Coral Gables, USA, November 1995.

[14] A. Pentland. Perceptual organization and the representation of natural form. *AI journal*, 28(2):1–38, 1986.

[15] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modelling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(7):715–729, July 1991.

[16] Richard Szeliski and David Tonnesen. Surface modeling with oriented particle systems. In Edwin E. Catmull, editor, *Computer Graphics (SIGGRAPH '92 Proceedings)*, volume 26, pages 185–194, July 1992.

[17] D. Terzopoulos and D. Metaxas. Dynamic 3-D models with local and global deformations : deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(7):703–714, July 1991.

[18] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: recovering 3-D shape and nonrigid motion. *AI Journal*, 36:91–123, 1988.

[19] Q.Z. Ye. The signed euclidean distance transform and its applications. In *International Conference on Pattern Recognition*, pages 495–499, 1988.

[20] A.L. Yuille, D.S. Cohen, and P.W. Hallinan. Feature extraction from faces using deformable templates. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, San Diego, June 1989.