M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 334
Submitted : International Symposium on Computer Vision 1995

Matching and Recognition Using Deformable Intensity Surfaces

Chahab Nastar

MIT – The Media Laboratory
Room E15-388, 20 Ames street
Cambridge, MA 02139
`nastar@media.mit.edu`

Alex Pentland

MIT – The Media Laboratory
Room E15-387, 20 Ames street
Cambridge, MA 02139
`sandy@media.mit.edu`

Abstract

We describe a novel approach for matching and recognition, based on deformable intensity surfaces. In this approach, the intensity surface of the image is modeled as a deformable thin plate, and its strain energy while deforming to a new image is measured for recognition. Each surface point has 3 degrees of freedom, thus capturing fine surface changes. The method is validated for recognition and interactive search in a large-scale face database.

1 Introduction

In this article, we introduce *deformable intensity surfaces* for object matching and recognition. Our method allows to deal effectively with object appearance variations. The basic idea is to warp the intensity surface of the test image into the intensity surface of the probe image by using a physics-based model, and then measure *the low-order strain energy* of the deformable surface for recognition. This recognition metric has to do with the *eigenmodes* of the intensity surface modeled as a thin-plate [18, 17].

We choose face recognition as an application for our method. For over 20 years, automatic face recognition has been the center of attention of computer vision researchers. Face recognition is indeed a difficult object recognition problem, because of the wide variation in the appearance of a specific face due to changes in pose, lighting, and facial expression. The applications are numerous, e.g. human-computer interaction, smart surveillance and image retrieval in a database (see [5] for a survey).

While earlier geometrical representations of faces focus on spatial distributions of facial features [10], photometric (i.e. image-based) methods have recently been implemented, including iso-density maps [16] or natural basis functions [22] ; view-based methods such as template matching [4, 2] and principal components analysis [27, 20, 12] have been particularly successful.

For most applications, there is need to develop a face-recognition system with the ability to perform recognition allowing changes in expression, lighting, and pose. Our system integrates these abilities and produced a recognition rate of 97% in a database of 7,562 faces, a significant improvement over previously reported results [20].

2 The intensity surface

2.1 The XYI space

Our idea of using intensity surfaces for matching and recognition comes from the observation that the transformation shape to intensity is quasi-linear under controlled lighting conditions ; in other terms, *the intensity of the 2D image reflects the actual 3D shape.* This essential observation is the basis of all shape from shading methods [8] ; however, unlike those methods, our aim is not actually to reconstruct depth information from a single 2D projection, but rather keep in mind that, under controlled lighting conditions, the changes in the image intensities from one image to the other reflect changes in their actual 3D shape. Mathematically, supposing the object of interest to be a Lambertian (or matte) surface, the amount of intensity reflected when illuminated by a single light source placed at infinity, is isotropic :

$$I(x, y) = \alpha \, \vec{N}(x, y).\vec{L} \tag{1}$$

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 :

$$(x, y) \longrightarrow I(x, y) \tag{2}$$

Our system focuses on recognition in the 3D space defined by $(x, y, I(x, y))$, that we will call the XYI space.

2.2 Resolution in I

In the XYI space, a 2D greyscale image of size $a.b$ is represented by a 3D binary image of size $a.b.g$, where

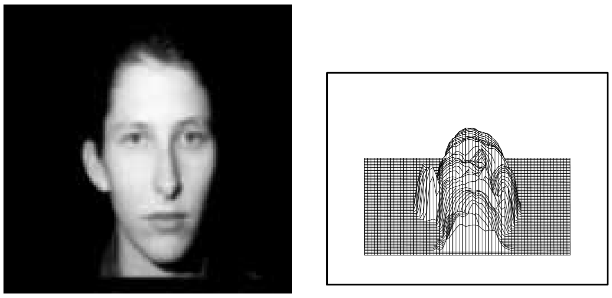

Figure 1: An image and its XYI surface representation

g is the number of greyscales. It is then clear that the value of g is an important parameter that we can modify for faster and better results.

For instance, in our test database of face images we have a resolution of $a = b = 128$, including the background. The face of the person itself has a typical size of 50×80 pixels. Therefore, the usual value of $g = 256$ is too high. We need to reduce g for several reasons :

- Reduce the size of the 3D binary image, and thus, computational time,
- Provide a smoothing of the intensity surface,
- Have the 3D binary image reflect as linearly as possible the actual 3D shape.

In our experiments, the value of $g = 16$ was chosen for faster computations (figure 2).

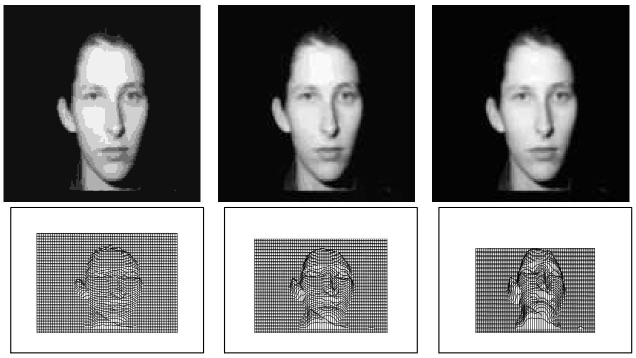

Figure 2: Different greylevels of the same image, and the corresponding XYI surface representations ($g = 8$, $g = 16$, $g = 32$)

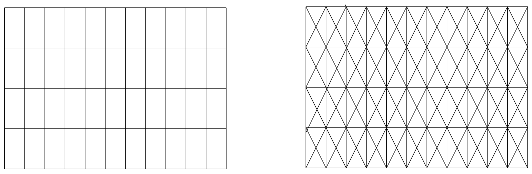

Figure 3: Surface meshes

3 Deformable intensity surfaces

In this section we define deformable intensity surfaces for image warping.

3.1 Theoretical background

Following the theory of active contour models [11, 26], several models have been developed that deal explicitly with deformable surfaces, among them : deformable superquadrics [19, 25], surface snakes [6, 13], particle systems [24], splines [3] and elastic thin plates [21, 18]. 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* [29]. Hence, deformable intensity surfaces is a new approach to matching and recognition.

The mathematical approach to our model is inspired by the one described in [18]. The intensity surface is modeled as a deformable mesh of size $N \times N'$ nodes (figure 3). It is ruled by Lagrangian dynamics [1] :

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}(t) \qquad (3)$$

where $\mathbf{U}$ 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 image force. Three equations of order $N \times N'$ of the above type coresponding to the three directions X, Y, I have to be solved. We can adapt the effective external force investigated in [18] to our problem : at each node M_i of the mesh points, the image force points to *the closest point P_i in the 3D binary image* discussed in section 2.2 :

$$\mathbf{F}(t) = [\ldots, \overrightarrow{M_i P_i}(t), \ldots]^T \qquad (4)$$

Euclidean distance algorithms can help us extract this force in each voxel of the 3D image, as a pre-processing [7, 28].

Note that our formulation provides an interesting alternative to optical flow methods, without the classical *brightness constraint* [9]. Indeed, the brightness constraint corresponds to a particular case of our formulation where the closest point P_i has to have the same intensity as M_i ($\overrightarrow{M_i P_i}$ is parallel to the XY plane). We do not make that assumption here.

3.2 Warping

In our formulation, deforming the intensity surface of *image*1 into the one of *image*2 takes place in 5 steps :

1. Reduce, if necessary, the number of greylevels in *image*1 and *image*2 down to the same number g of greylevels.

2. Initialize the deformable surface S as a subsampling of the intensity surface of *image*1.

3. Convert *image*2 to its 3D binary representation, *image*3.

4. Compute distance maps at each voxel of *image*3.

5. Let S deform dynamically in *image*3 (equation (3)) with the external force derived from the distance maps created at step 2.

Note that steps 1 to 4 are pre-processing steps. Steps 1 and 2 provide respectively intensity and spatial smoothing of the image, an essential prerequisite before actually considering $I(x, y)$ as a surface.

3.2.1 An example

Figure 4 illustrates four warping experiments. The first example is the warping of two different faces. The three other outline facial expression, changes in lighting, and out-of-image plane rotation.

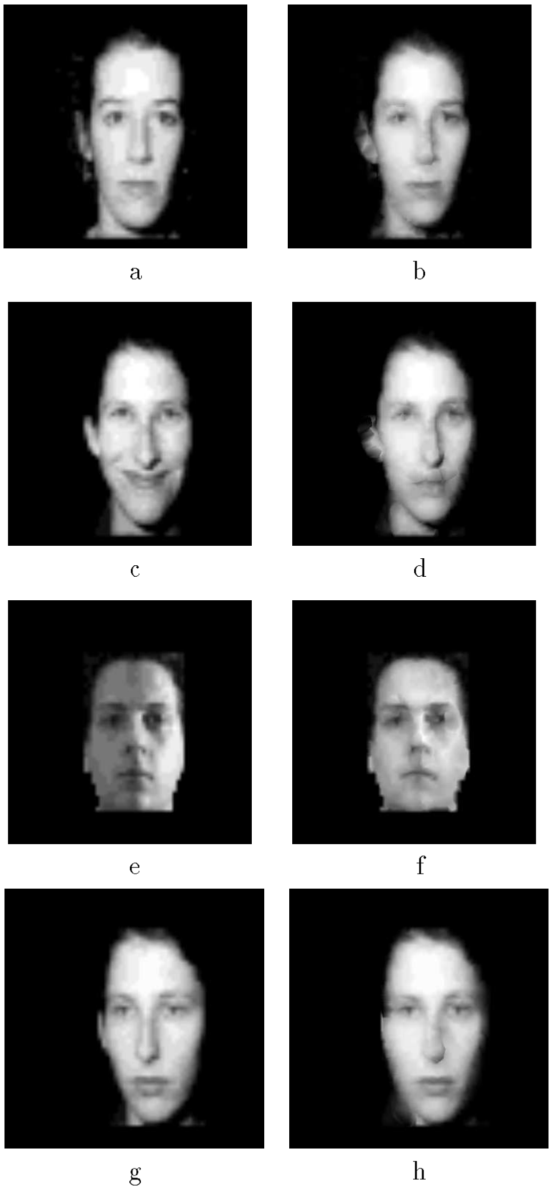

Figure 4: Warping face images in XYI space. **a. b.** Warping two different faces. **c. d.** Warping smiling face to neutral face. **e. f.** Warping face with transversal light into face with head-on light. **g. h.** Warping rotated face to neutral face.

4 Vibration modes of the surface

4.1 Theory

The general applicability of eigenvector decomposition methods for appearance-based 3D object recognition has been convincingly demonstrated in [15]. For deformable surfaces, a powerful eigenvector decomposition is *modal analysis* [21, 18, 17].

The vibration modes $\boldsymbol{\phi}$ of the previous deformable surface are the vector solutions of the eigenproblem :

$$\mathbf{K}\boldsymbol{\phi} = \omega^2 \mathbf{M}\boldsymbol{\phi} \tag{5}$$

where $\omega(i)$ is the i-th eigenfrequency of the system. Solving the governing equations in the modal basis leads to scalar equations where the unknown $\tilde{u}(i)$ is the amplitude of mode i [1] :

$$\ddot{\tilde{u}}(i) + \tilde{c}_i \dot{\tilde{u}}(i) + \omega(i)^2 \tilde{u}(i) = \tilde{f}_i(t) \tag{6}$$

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

$$\mathbf{U} \approx \sum_{p=1}^{P} \sum_{p'=1}^{P'} \tilde{u}(p, p') \boldsymbol{\phi}(p, p') \tag{7}$$

with $P \times P' \ll N \times N'$. The modal superposition equation (7) can be seen as a Fourier expansion with high-frequencies neglected, as outlined in [17].

We make use of *the analytic expressions of the modes* which are known sine and cosine functions for specific surface topologies. For quadrilateral surface meshes that have plane topology (which is the case of the intensity surfaces), the eigenfrequencies are [17] :

$$\omega^2(p, p') = 4K/M\,(\sin^2 \frac{p\pi}{2N} + \sin^2 \frac{p'\pi}{2N'}) \tag{8}$$

K is the stiffness of each spring, M the mass of each node, p and p' are the mode parameters. The modes of vibration are :

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

where $n = 1, \ldots, N$ and $n' = 1, \ldots, N'$. These analytic expressions avoid the call to costly eigenvector-extraction routines ; moreover, they allow the total number of modes $P \times P'$ to be easily adjusted.

4.2 Strain energy as a similarity measure

The strain energy describing the amount of energy needed to deform a surface into another is by definition [1, 23] :

$$E_{strain} = \frac{1}{2} \mathbf{U}^T \mathbf{K} \mathbf{U}$$

This expression is much simpler when developed in the modal (or vibration) basis of the system :

$$E_{strain} = \frac{1}{2} \sum_{p=1}^{N} \sum_{p'=1}^{N'} \omega^2(p, p')\, \tilde{u}^2(p, p')$$

Therefore, the low-order strain energy corresponding to $P \times P'$ low-order modes is :

$$E_{strain}(P, P') = \frac{1}{2} \sum_{p=1}^{P} \sum_{p'=1}^{P'} \omega^2(p, p')\, \tilde{u}^2(p, p') \tag{10}$$

with $P \times P' \leq N \times N'$.

4.2.1 Application to face images

Changes in facial expression are usually high-frequency phenomenons. The global shape of the face is indeed unchanged while some specific features of the face move or change shape. On the other hand, a small variation in pose or changes in lighting will excite low-frequency modes.

Thus, the low-order strain energy $E_{strain}(P, P')$ is an efficient similarity measure. The justification of its use for robust face recognition relies on the following observations (see figure 5) :

- Changes in expression do not affect the low-order strain energy since high frequencies are cut off (summation is over a set of low-frequency modes).

- Small motions or changes in lighting have only small effect on the strain energy since low-frequency strain energies are negligible (low ω^2).

4.3 How many modes ?

The number of modes to be superimposed is one of the main problems of any vibration-based technique. Indeed, the optimal number cannot be adjusted *a priori* for any application ; it clearly depends on the frequency components of the motion. Basically, the optimal number of modes will be *as small as possible*, but should capture *a representative sketch* of the surface deformation. A typical *a priori* value covering many types of standard deformations is the quarter of the degrees of freedom in the system [1]. Therefore we set :

$$P = N/4 \text{ and } P' = N'/4$$

Note that these values correspond to a compression factor equal to 16.

Consider now the deformation occuring in figure 4(a and b). The low-frequency recovery of the deformation is displayed in figure 6.

5 Face recognition in a database

The Media Lab database contains 7,562 "mugshot" face images of approximately 3,000 people. Each person has at least 2 images in the database. The faces are roughly aligned in the same position in all images, so that no normalization against translation, rotation or scale is necessary ; however, changes in lighting and facial expressions, as well as out-of-plane rotations frequently occur.

We select a reference face in among the 7,562 images in the Media Lab database. We then warp all other faces of the database to the reference face, compute the strain

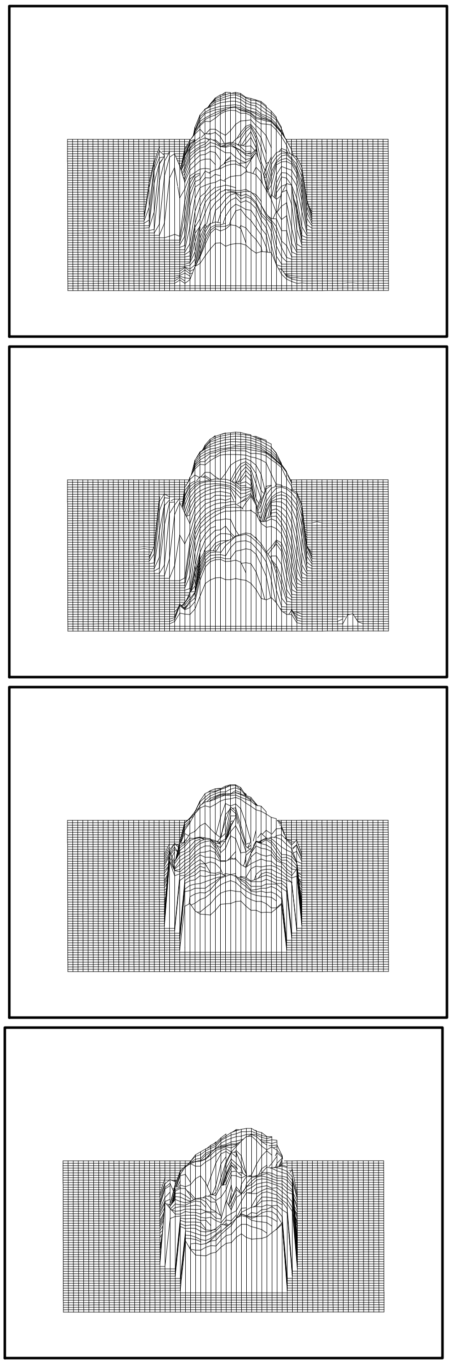

Figure 5: Face intensity surfaces. From top to bottom : Neutral Claire. Smiling Claire. Joel with head-on light. Joel with oblique light.

Figure 6: Low frequency modal recovery of the deformation in figure 4(a and b). Compression is 16.

energies and store them as a set of coefficients. The i-th face is stored as : $\tilde{u}_i(p,p')$. The relative distance between the faces labeled i and j is :

$$d(i,j) = \frac{1}{2}\sum_{p=1}^{P}\sum_{p'=1}^{P'} \omega^2(p,p')\left(\tilde{u}_i(p,p') - \tilde{u}_j(p,p')\right)^2 \quad (11)$$

To assess the recognition rate, we use a *nearest neighbor rule*. If the most similar face was of the same person then a correct recognition was scored. In our experiment, the system produced a recognition accuracy of 97%, which corresponds to only six mistakes in matching views of 200 people randomly chosen in the database of 7,562 images . This is significantly better than had previously been reported for this dataset [20]. Examples of recognition are shown in figure 7.

6 Conclusion and future work

We introduce deformable intensity surfaces for image matching and object recognition. The low-order strain energy, i.e. the strain energy with high frequency components removed, is an efficient metric for object recognition when the object appearance varies. Our model has 3 degrees of freedom (X,Y and I), which makes it a more powerful approach than single degree of freedom techniques [27].

For face images, our approach allows expression- lighting- and pose-insensitive recognition in a face database. We present results of recognition in the Media Lab database of 7,562 faces, with an accuracy of 97%, a significant improvement over previously reported techniques.

Future work will take advantage of feature recognition [20, 14]. Incorporating feature correspondences as constraints in our system will make both the matching and the recognition steps more robust, thus dealing more efficiently with complex object appearance variations (figure 8).

probe image best match second match third match

Figure 7: Recognition examples. In each row, we search for the best match using the far left face as a probe into the database. Only the 3 best matches are displayed in each case. Similarity decreases left to right.

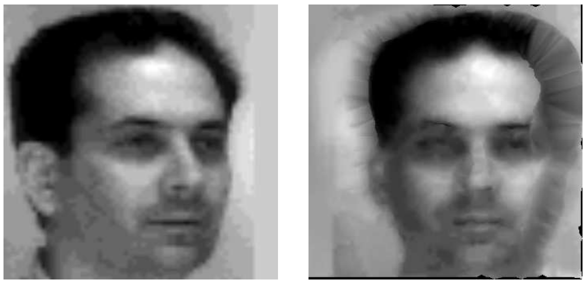

Figure 8: Warping rotated face to frontal face in XYI. Feature constraints are incorporated to the intensity surface warping.

7 References

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

[2] D. J. Beymer. Face recognition under varying pose. Technical Report AI memo 1461, MIT, December 1993.

[3] 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.

[4] R. Brunelli and T. Poggio. Face recognition : Features vs. templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10), October 1993.

[5] R. Chellappa, S Sirohey, C.L. Wilson, and C.S. Barnes. Human and machine recognition of faces : A survey. Technical Report CAR-TR-731, CS-TR-3339, University of Maryland, August 1994.

[6] I. Cohen, L. D. Cohen, and N. Ayache. Using deformable surfaces to segment 3-D images and infer differential structures. *Computer Vision, Graphics, and Image Processing*, 1992.

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

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

[9] B.K.P. Horn and G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[10] T. Kanade. *Computer Recognition of Human Faces.* Basel and Stuttgart : Birkhauser, 1977.

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

[12] M Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), January 1990.

[13] 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*, pages 518–523, Berlin, June 1993. IEEE.

[14] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE Proceedings of the Fifth International Conference on Computer Vision*, Cambridge, USA, june 1995.

[15] H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14(5), 1995.

[16] O. Nakamura, S Mathur, and T. Minami. Identification of human faces based on iso-density maps. *Pattern Recognition*, 24(3), 1991.

[17] 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.

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

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

[20] A. Pentland, B. Moghaddam, T. Starner, and M. Turk. View based and modular eigenspaces for face recognition. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, 1994.

[21] 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.

[22] R. P. N. Rao and D. H. Ballard. Natural basis functions and topographic memory for face recognition. In *IJCAI*, 1995.

[23] S. Sclaroff and A. Pentland. A modal framework for correspondence and description. In *Proceedings of the Fourth International Conference on Computer Vision (ICCV '93)*, Berlin, May 1993.

[24] 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.

[25] 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.

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

[27] M. Turk and A. Pentland. Face recognition using eigenfaces. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pages 586–591, 1991.

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

[29] 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.