M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 368
Appears in: *Fourth European Conference on Computer Vision, Cambridge, UK, April 1996.*

Generalized Image Matching: Statistical Learning of Physically-Based Deformations

Chahab Nastar, Baback Moghaddam and Alex Pentland
Perceptual Computing Section, The Media Laboratory,
Massachusetts Institute of Technology
20 Ames Street, Cambridge MA 02139, U.S.A.

Abstract

We describe a novel approach for image matching based on deformable intensity surfaces. In this approach, the intensity surface of the image is modeled as a deformable 3D mesh in the $(x, y, I(x, y))$ space. Each surface point has 3 degrees of freedom, thus capturing fine surface changes. A set of representative deformations within a class of objects (e.g. faces) are statistically learned through a Principal Components Analysis, thus providing a priori knowledge about object-specific deformations. We demonstrate the power of the approach by examples such as image matching and interpolation of missing data. Moreover this approach dramatically reduces the computational cost of solving the governing equation for the physically based system by approximately three orders of magnitude.

1 Introduction

In recent years, computer vision research has witnessed a growing interest in eigenvector analysis and subspace decomposition methods [15]. This general analysis framework lends itself to several closely related formulations in object modeling and recognition which employ the *principal modes* or characteristic *degrees-of-freedom* for description. The identification and parametric representation of a system in terms of these principal modes is at the core of recent advances in physically-based modeling [20, 17] and parametric descriptions of shape [7, 2, 11]. On the other hand, *view-based* eigentechniques have recently provided some of the best results in object recognition [21, 19].

In this paper, we propose a new method which combines both the physically-based modes of vibration and the statistically-based modes of variation. In view of some recent critiques of physical modeling (e.g. [4]) our motivation here is to ground physically-based models in actual real-world statistics in order to obtain a more realistic and data-driven model for the underlying phenomenon [13, 6].

Furthermore, we seek to unify the shape and texture components of an image in a single compact mathematical framework. Current work in the area of image-based object modeling deals with the shape (2D) and texture (grayscale) components of an image in an independant manner [3, 5]. Our novel representation combines both the spatial (XY) and grayscale (I) components of the image into a 3D surface (or manifold) and then efficiently solves for a dense correspondance map in the XYI space. This "manifold matching" technique can be viewed as a more general formulation for image correspondance which, unlike optical flow, does *not* require a constant brightness assumption.

In principal, any two image manifolds can be matched in this way (though sometimes erroneously), therefore we must further constrain the space of allowable manifold deformations to specific object classes (eg., frontal views of faces). These characteristic deformations (or "principal warps") are learned through a statistical Principal Components Analysis (PCA) [10] which identifies the principal subspace in which the final correspondance field must lie. Since the Karhunen-Loeve Transform (KLT) [12] in PCA corresponds to a unitary linear change of basis, which can be appended to the modal transform used in solving the physical system, we can ultimately derive a compact reduced-order form of the governing equation which combines both the dynamics of the physical system and the "learned" deformations which were observed in actual training data.

2 Deformable Intensity Surfaces for Image Matching

Our 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 terms, *the intensity of the 2D image reflects the actual 3D shape.* Our system focuses on matching and recognition in the 3D space defined by $(x, y, I(x, y))$, that we will call the XYI space (see [18] for details).

In our formulation, deforming the intensity surface of $image1$ into the one of $image2$ in XYI takes place in 5 steps :

1. Reduce, if necessary, the number of graylevels in $image1$ and $image2$ down to the same number g of graylevels (typically $g = 32$).
2. Initialize the deformable surface S as a subsampling of the intensity surface of $image1$.
3. Convert $image2$ to its 3D binary representation, $image3$.
4. Compute Euclidean distance maps at each voxel of $image3$ [8, 22].
5. Let S deform dynamically in $image3$ 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

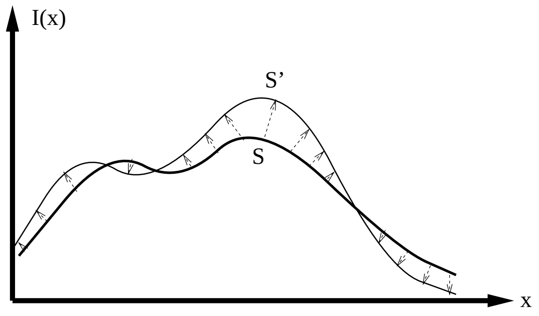

Figure 1: Intensity surface S being pulled towards S' by image forces

of the image. The dynamic process of step 5 is described in [17] ; to sum up, the intensity surface S is modeled as a deformable mesh of size $N = n \times n'$ nodes, ruled by Lagrangian dynamics :

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}(t) \tag{1}$$

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 (or "image") force. The above equation is of order $3N$.

At each node M_i of the mesh, the image force points to *the closest point* P_i in the 3D binary image $image3$ [17]. Figure 1 shows a representation of the deformation process. Note that the external forces (dashed arrows) *do not* necessarily correspond to the final displacement field of the surface since the closest point P_i is updated at each time iteration. The elasticity of the surface provides an intrinsic smoothness constraint for computing the final displacement field.

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[1] 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.

The vibration modes $\phi(i)$ of the previous deformable surface are the vector solutions of the eigenproblem [1] :

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

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 :

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

The closed-form expression of the displacement field is now :

$$\mathbf{U} \approx \sum_{i=1}^{P} \tilde{u}(i)\phi(i) \tag{4}$$

[1]In fact, by simply disabling the I component of our deformations we can obtain a standard 2D deformable mesh which yields correspondances similar to an optical flow technique with thin-plate regularizers.

with $P \ll 3N$, which means that only P scalar equations of the type of (3) need to be solved. The modal superposition equation (4) can be seen as a Fourier expansion with high-frequencies neglected [16].

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 of the system are [16] :

$$\omega^2(p, p') = 4K/M\,(\sin^2 \frac{p\pi}{2n} + \sin^2 \frac{p'\pi}{2n'}) \tag{5}$$

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 :

$$\phi(p, p') = [\ldots, \cos \frac{p\pi(2i-1)}{2n} \cos \frac{p'\pi(2j-1)}{2n'}, \ldots]^T \tag{6}$$

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

3 Statistical Modeling

In theory, our deformable intensity surface can undergo any possible deformation. Thus, it seems interesting to *learn* the deformations of a specific class of objects and add them as *constraints* into our system. This is an important step for guiding the deformations of our mesh when performed within a specific object class and also allows us to deal with occlusions and missing data, as we shall see later.

Our approach to learning the space of allowable manifold deformations particular to a specific object class Ω (eg., frontal faces) is that of *unsupervised learning*. Particularly, we perform a PCA on a selected training set of deformations in order to recover the principal components of the warps. This approach is actually part of a more complete statistical formulation for estimating the probability density function of these warps in the high-dimensional vector space $\tilde{\mathbf{U}} \in \mathcal{R}^P$ (see [14]). The estimated class-conditional density $P(\tilde{\mathbf{U}}|\mathbf{\Omega})$ can be ultimately used in a Bayesian framework for a variety of tasks such as regression, interpolation, inference and classification. However, in this paper, we have concentrated mainly on the dimensionality-reduction aspect of PCA in order to obtain a lower-dimensional subspace in which to solve for the manifold correspondance field.

Given a training set of suitable warp vectors $\{\tilde{\mathbf{U}}^t\}$ for $t = 1 \ldots N_T$, the principal warps are obtained by solving the eigenvalue problem

$$\Lambda \;=\; \mathbf{E}^T \Sigma \mathbf{E} \tag{7}$$

where Σ is the covariance matrix of the training set, $\mathbf{E}$ is the eigenvector matrix of Σ and Λ is the corresponding diagonal matrix of eigenvalues. The unitary matrix $\mathbf{E}$ defines a coordinate transform (rotation) which *decorrelates* the data and makes explicit the *invariant subspaces* of the matrix operator Σ. In PCA, a partial KLT is performed to identify the largest-eigenvalue eigenvectors and obtain a principal component feature vector $\hat{\mathbf{U}} = \mathbf{E}_{\mathrm{M}}^T\,(\tilde{\mathbf{U}} - \tilde{\mathbf{U}}_0)$, where $\tilde{\mathbf{U}}_0$ the mean warp vector and $\mathbf{E}_M$ is a submatrix of $\mathbf{E}$ containing the principal eigenvectors. This KLT can be seen as a linear transformation $\hat{\mathbf{U}} = \mathcal{T}(\tilde{\mathbf{U}}) : \mathcal{R}^P \rightarrow \mathcal{R}^L$ which extracts a lower-dimensional subspace of the KL basis corresponding to the maximal eigenvalues. These principal components preserve the major linear correlations in the data and discard the minor ones.[2]

By ranking the eigenvectors of the KL expansion with respect to their eigenvalues and selecting the first L principal components we form an orthogonal decomposition of the vector space $\mathcal{R}^P$ into two mutually exclusive and complementary subspaces: the principal subspace (or feature space) $\{\mathbf{E}_i\}_{i=1}^{L}$ containing the principal components and its orthogonal complement $\bar{F} = \{\mathbf{E}_i\}_{i=L+1}^{P}$. In this paper, we simply discard the orthogonal subspace and work entirely within the principal subspace $\{\mathbf{E}_i\}_{i=1}^{L}$, hereafter referred to simply by the matrix $\mathbf{E}$.

4 Combining Physics and Statistics

Instead of solving the unconstrained governing equation (1), we compute the projection of the unknown $\mathbf{U}$ (dimension : 3N = 3nn'), first into a modal sub-basis (dimension P), then into a KL subspace (dimension L) :

$$\mathbf{U} \xrightarrow{\boldsymbol{\Phi}} \tilde{\mathbf{U}} \xrightarrow{\mathbf{E}} \hat{\mathbf{U}} \quad (8)$$

The first transform is the projection into the modal subspace :

$$\mathbf{U} = \boldsymbol{\Phi}\tilde{\mathbf{U}} \quad (9)$$

The second transform is the projection of the modal amplitudes into the PCA subspace :

$$\tilde{\mathbf{U}} = \mathbf{E}\hat{\mathbf{U}} + \tilde{\mathbf{U}}_0 \quad (10)$$

Equations (9) and (10) yield the global transform :

$$\mathbf{U} = \boldsymbol{\Psi}\hat{\mathbf{U}} + \mathbf{U}_0 \quad (11)$$

where the global transformation matrix Ψ is simply : $\boldsymbol{\Psi} = \boldsymbol{\Phi}\mathbf{E}$ and $\mathbf{U}_0 = \boldsymbol{\Phi}\tilde{\mathbf{U}}_0$. Note that Ψ is a rectangular orthogonal matrix.

By premultiplying equation (1) by Ψ^T and changing unknowns (equation (11)), we obtain :

$$\boldsymbol{\Psi}^T \mathbf{M}\, \boldsymbol{\Psi}\, \ddot{\hat{\mathbf{U}}} + \boldsymbol{\Psi}^T \mathbf{C}\, \boldsymbol{\Psi}\, \dot{\hat{\mathbf{U}}} + \boldsymbol{\Psi}^T \mathbf{K}\, \boldsymbol{\Psi}\, \hat{\mathbf{U}} = \boldsymbol{\Psi}^T \mathbf{F}(t) - \boldsymbol{\Psi}^T \mathbf{K}\, \mathbf{U}_0 \quad (12)$$

Let :

$$\hat{\mathbf{M}} = \boldsymbol{\Psi}^T \mathbf{M} \boldsymbol{\Psi} \quad (13)$$

$$\hat{\mathbf{C}} = \boldsymbol{\Psi}^T \mathbf{C} \boldsymbol{\Psi} \quad (14)$$

$$\hat{\mathbf{K}} = \boldsymbol{\Psi}^T \mathbf{K} \boldsymbol{\Psi} = \mathbf{E}^T \boldsymbol{\Omega}^2 \mathbf{E} \quad (15)$$

$$\hat{\mathbf{F}}(t) = \boldsymbol{\Psi}^T \mathbf{F}(t) - \boldsymbol{\Psi}^T \mathbf{K} \mathbf{U}_0 = \boldsymbol{\Psi}^T \mathbf{F}(t) - \mathbf{E}^T \boldsymbol{\Omega}^2 \tilde{\mathbf{U}}_0 \quad (16)$$

Note that the new mass, damping and stiffness matrices, as well as the new external force, do not involve heavy computations because : *(i)* we make the common assumption that $\mathbf{M}$ and $\mathbf{C}$ are scalar matrices ($\mathbf{M} = M\mathbf{I}$, $\mathbf{C} = C\mathbf{I}$ where M and C are mass and damping scalars) , and *(ii)* Ω^2 is a diagonal matrix. We now end up with the standard Lagrangian equation of unknown $\hat{U}$.

$$\hat{\mathbf{M}}\ddot{\hat{\mathbf{U}}} + \hat{\mathbf{C}}\dot{\hat{\mathbf{U}}} + \hat{\mathbf{K}}\hat{\mathbf{U}} = \hat{\mathbf{F}}(t) \quad (17)$$

Solving this equation for $\hat{\mathbf{U}}$ and then changing basis back to the canonical basis (equation (11)) provides the estimated displacement $\mathbf{U}$. By using this method, the resulting displacement $\mathbf{U}$ is constrained to lie along those learned deformation modes that are characteristic of the object class.

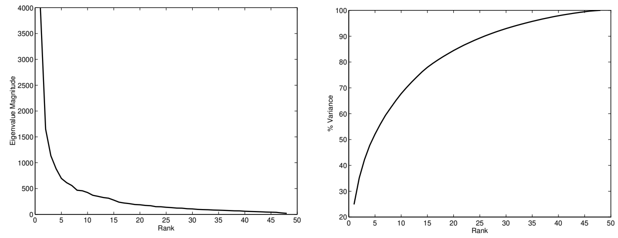

Figure 3: Left : Eigenvalue spectrum of the PCA transform. Right : Cumulative eigenvalue spectrum of the PCA transform

5 Experimental Results

We conduct our experiments with facial imagery. The manifold matching technique described in this paper requires rough alignment of the two input images in order to function properly. In our experiments, this alignment was obtained using an automatic face-processing system which extracts faces from the input image and normalizes for translation, scale and slight rotations (both in-plane and out-of-plane). This system is described in detail in [14].

For the learning phase of our technique, we choose a set of 50 faces to be warped into a reference face. Each of these faces has a $N = 128 \times 128$ resolution, and the manifolds are matched in a modal subspace whose dimension is suitably chosen $P = 3 \times 128^2/4^2 = 3072$ [18]. We then perform a Principal Components Analysis on the spectra of these warps.

Figure 2 shows the modes of variation along individual KL-eigenvectors extracted from the learning set. For example, we can see that $\overrightarrow{E_1}$ represents change in global headshape (as well as the size of the eyes). Eigenvectors $\overrightarrow{E_2}$ and $\overrightarrow{E_3}$ represent a change in the chin size and forehead, respectively. Higher-order eigenvectors, for example $\overrightarrow{E_{10}}$ represent subtler variations in facial appearance (e.g. eye shape).

By looking at the KL-eigenvalues, it is easy to draw the percentage of the data variance that is captured versus the number of eigenvalues. Figure 3 shows that 90% of the data is adequately captured by $L = 25$ principal eigenvectors.

5.1 Subspace Warps

Figure 4 shows an example of matching a test image to that of the reference using both the unconstrained and

[2] In practice the number of training images N_T is far less than the dimensionality of the data, P, consequently the covariance matrix Σ is singular. However, the first $L < N_T$ eigenvectors can always be computed (estimated) from N_T samples using, for example, a Singular Value Decomposition.

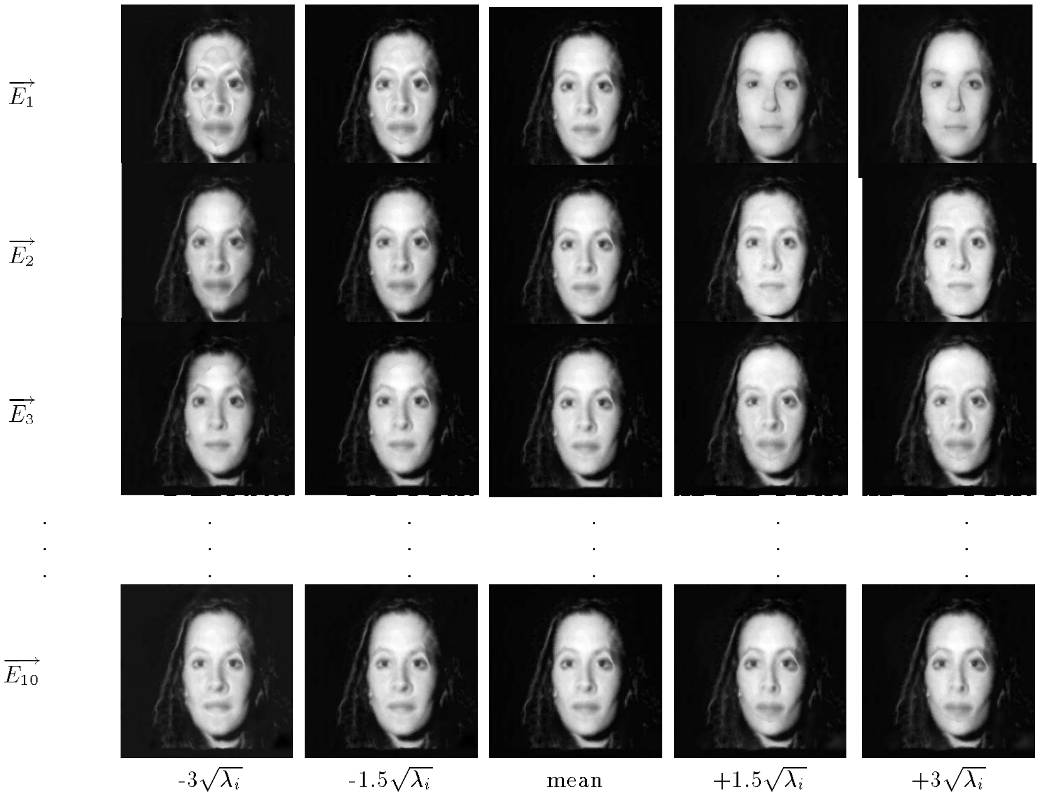

Figure 2: Modes of variation of the manifold

Figure 4: From left to right : test image ; reference image ; unconstrained warp in modal space ; constrained warp in KL-space

constrained warps. This basic example illustrates how a dense correspondence field can be obtained between two images from different objects.

Figure 5 displays the modal spectrum and its reconstruction in the KL space. The total reconstruction error is on the order of 4%, demonstrating that by solving the reduced-order physical system (equation (12)), we have not significantly sacrificed accuracy. In addition, solving this equation requires considerably less computation. The degrees of freedom in the original mesh were $3N = 3 \times 128 \times 128 \approx 50,000$. In the modal subspace, the degrees of freedom were reduced to $P = 3 \times 32 \times 32 \approx 3,000$, and finally in the KL subspace, the degrees of freedom were further reduced to $L = 25$, thus achieving a compression factor of approximately $5000 : 1$.

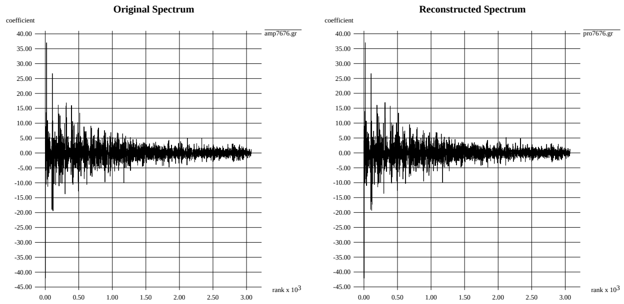

Figure 5: Left : the original spectrum of the deformation. Right : reconstruction of the spectrum in the KL-subspace.

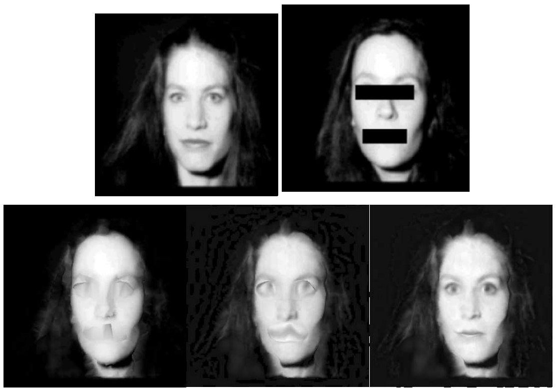

Figure 6: **Top** : we wish to warp the left image into the right image **Bottom** : image warps ; left : in the real space. center : in the unconstrained modal subspace. right : in the constrained principal subspace

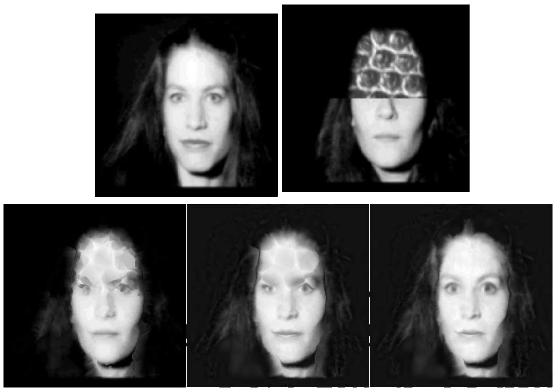

Figure 7: See caption of figure 6

5.2 Interpolation of Missing data

One of the advantages of learned warps is that, during the matching process, the deformations are constrained for a specific object. Consequently, invalid deformations arising out of missing data (e.g. object occlusion) are automatically disallowed.

The first example illustrates an experiment where regions of the face were occluded with a black bar (to simulate occlusion or incomplete data), as shown in figure 6 (top row). If we attempt an unconstrained warp in the modal space, an invalid reconstruction will be obtained (figure 6 bottom left and center). On the other hand, if the deformation is constrained by the learned modes, we obtain a better reconstruction of the missing data as shown in figure 6 (bottom right). This example illustrates how our principal warp formulation effectively functions as a model-based image interpolant for a given class of objects.

The second example is similar in spirit to the first, except where the missing data is replaced by an arbitrary image region (in this case a texture), for example when one object partially occludes another. Here once again we see how the learned principal warps can yield a much better reconstruction and interpolation of non-matching image regions (figures 7).

6 Conclusions

We have described a novel approach for image matching based on deformable intensity surfaces. In this approach, the intensity surface of the image is modeled as a deformable surface embedded in XYI space. Our approach is thus a *generalization* of optical flow and deformable shape matching methods (which consider only changes in XY), of statistical texture models such as "eigenfaces" (which consider only changes in I an assume an already existing XY correspondance), and of hybrid methods which treat shape and texture separately and sequentially.

We have further shown how to tailor the space of allowable XYI deformations to fit the actual variation found in individual target classes. This was accomplished by a statistical analysis of observed image-to-image deformations using a Principal Components Analysis. The result is that the image deformation is restricted to the subspace of physically-plausible deformations. In the process, the dimensionality of the matching and the numerical complexity of the governing equation are drastically reduced.

By considering only the low-dimensional subspace of plausible deformations, we make the image matching process more robust and more efficient. We in effect "build in" statistical *a priori* knowledge about how the object can vary in order to obtain the best image-to-image match possible. To illustrate the power of this method we have shown that we can interpolate missing data despite occlusions and noise, and that we can use this method to obtain very compact image descriptions.

References

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

[2] A. Baumberg and D. Hogg. Learning flexible models from image sequences. In *Proceedings of the Third European Conference on Computer Vision 1994 (ECCV'94)*, Stockholm, Sweden, May 1994.

[3] David Beymer. Vectorizing face images by interleaving shape and texture computations. A.I. Memo No. 1537, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1995.

[4] T. Boult. Physics in a fantasy world vs. robust statistical estimation. In *NSF/ARPA workshop on 3D Object Representation in Computer Vision*, New York, USA, December 1994.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[6] T.F. Cootes and C.J. Taylor. Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5), 1995.

[7] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Training models of shape from sets of examples.

In *Proceedings of the British Machine Vision Conference*, Leeds, 1992.

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

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

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

[11] C. Kervrann and F. Heitz. Learning structure and deformation modes of nonrigid objects in long image sequences. In *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition*, 1995.

[12] M.M. Loeve. *Probability Theory*. Van Nostrand, Princeton, 1955.

[13] J. Martin, A. Pentland, and R. Kikinis. Shape analysis of brain structures using physical and experimental modes. In *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, USA, June 1994.

[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] 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.

[17] 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.

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

[19] 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.

[20] 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.

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

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