M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 267
Appeared in *Proc. of the* 2^{nd} *IEEE CAD-Based Vision Workshop*, **Champion, PA, Feb. 1994 (pp. 258–265)**

Object Recognition and Categorization Using Modal Matching

Stan Sclaroff and Alex P. Pentland

Perceptual Computing Section, The Media Laboratory
Massachusetts Institute of Technology
20 Ames St., Cambridge MA 02139

Abstract

Recently we have proposed a new method for establishing correspondences and computing canonical descriptions [6], based on the idea of describing objects by their generalized symmetries, as defined by the object's vibration or deformation modes. In this paper we extend this method to the problems of object recognition and categorization. Examples of recognizing hand tools and airplanes are shown.

1 Introduction

Imagine that we are given two sets of image feature points, and our goal is to determine if they are from two similar objects. One way to solve this problem is to try to subject the first set of points to all possible translations and rotations in an attempt to get the first set of points to closely overlay the second set. If, however, the points were given in body-centered coordinates rather than Cartesian coordinates, the problem would be nearly trivial. This suggests that the best way to match two sets of feature points is to first compute a body-centered coordinate frame.

Our approach to this problem has three parts:

1. We compute a shape description that is robust with respect to sampling by using Galerkin interpolation, which is the mathematical underpinning of the finite element method.
2. We utilize a new type of Galerkin interpolant based on Gaussians that allows us to efficiently derive our shape parameterization directly from the data.
3. We then use the eigenmodes of this shape description to obtain a canonical, frequency-ordered orthogonal coordinate system. This coordinate system may be thought of as the shape's *generalized symmetry axes*.

By describing feature point locations in this body-centered coordinate system, it is easy to match corresponding points, and to measure the similarity of different objects. This allows us to recognize objects, and to determine if different objects are related by simple physical transformations.

A flow-chart of our method is shown in Figure 1. For each image we start with feature point locations and use these locations as nodes in building a finite element model of the shape. We construct a model of the shape by covering each feature point with a Gaussian blob of rubbery material; if there is segmentation information, then we can fill in interior areas and trim away material that extends outside of the shape.

We then compute the *modes of free vibration* of this model. These modes are the eigenvectors of the finite element model, and provide an orthogonal, frequency-ordered description of the shape and its natural deformations. The first three modes are the rigid body modes of translation and rotation, and the rest are nonrigid modes; in general, low-frequency modes describe global deformations, while higher-frequency modes do not. This frequency ordering of shape deformation will prove very useful for shape matching and comparison.

The modes also form an orthogonal *object-centered* coordinate system for describing feature locations. That is, each feature point location can be uniquely described in terms of *how it moves within each deformation mode*. In our technique, two groups of features are compared in the modal eigenspace; the important idea here is that the low-order vibration modes computed for two similar objects will be very similar — even in the presence of affine deformation, nonrigid deformation, local shape perturbation, or noise.

Finally, given correspondences between many of the feature points on two objects, we can measure their difference in shape. Because the modal framework decomposes deformations into an orthogonal set, we can selectively measure rigid-body differences, or low-order projective-like deformations, or deformations that are primarily local. Consequently, we can recognize objects in a very flexible and general manner.

In this paper we first give a brief review of our approach to computing correspondence, and then extend our approach to the problems of object recognition and category classification. We will demonstrate this new method on examples of recognizing and categorizing images of airplanes and hand tools.

2 Background and notation

Shape correspondence has previously been formulated as an equilibrium problem, which has the attractive feature of allowing integration of physical constraints [2; 3; 4; 5; 8; 10; 9]. To accomplish this, we first imagine that the collection of feature points in one image is attached by springs to an elastic body. Under the load exerted by these springs, the elastic body will deform to match the shape outlined by the set of feature points. If we repeat this procedure in each image, we can obtain a feature-to-feature correspondence by noting which points project to corresponding locations on the two elastic bodies.

Unfortunately, deformable models do not by themselves provide a method of computing canonical descriptions, and are therefore very difficult to use for shape comparison. To address this problem, we proposed a method in which shapes are represented as modal deformations from some prototype object [4; 5]. By describing deformation in terms of the eigenvectors of the prototype object's finite element stiffness matrix, we were able to obtain a robust, frequency-ordered shape description. In addition, high-frequency eigenvectors can be discarded to obtain overconstrained, canonical descriptions of the equilibrium solution.

Cootes, *et al.* have also proposed a method for obtaining canonical descriptions. They parameterize a deformable model based on a snake's principal variations over a given training set [1]. In the Cootes method, we are limited to measuring only the principal deformations revealed in the training phase, and therefore cannot easily build general models for recognition.

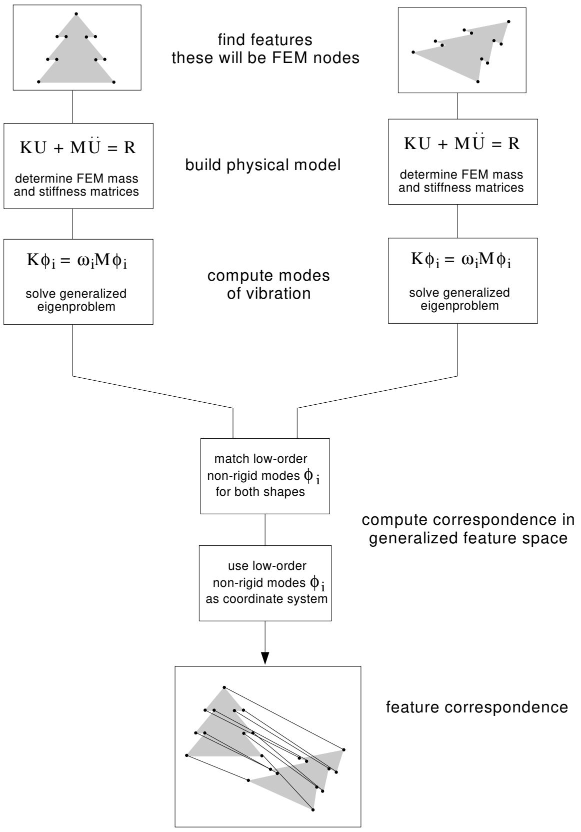


Figure 1: System diagram.

Both approaches are limited by the requirement that every object be described as the deformations of a *single* prototype object. This has the effect of implicitly imposing an *a priori* parameterization upon the sensor data, and therefore implicitly determining the correspondences between data and the prototype. Recently, therefore, we have generalized our finite element formulation so that we can obtain the modal shape invariants directly from the sensor data [6]. This has allowed us to both solve correspondence problems and compute robust, canonical descriptions for data of any dimensionality. The remainder of this section will provide a brief overview of our approach.

2.1 Finite element method

The most common numerical approach for solving equilibrium problems of this sort is the *finite element method*. The major advantage of the finite element method is that it uses the Galerkin method of surface interpolation. This provides an analytic characterization of shape and elastic properties over the whole surface, rather than just at the nodes (nodes are typically the spring attachment points).In Galerkin's method, we set up a system of polynomial shape functions that relate the displacement of a single point to the relative displacements of all the other nodes of an object. By using these functions, we can calculate the deformations which spread uniformly over the body as a function of its constitutive parameters.

In general, the polynomial shape function for each element is written in vector form as:

$$\mathbf{u}(\mathbf{x}) = \mathbf{H}(\mathbf{x})\mathbf{U} \tag{1}$$

where $\mathbf{H}$ is the interpolation matrix, $\mathbf{x}$ is the local coordinate of a point in the element where we want to know the displacement, and $\mathbf{U}$ denotes a vector of displacement components at each element node.

For most applications it is necessary to calculate the strain due to deformation. Strain ϵ is defined as the ratio of displacement to the actual length, or simply the ratio of the change in length. The polynomial shape functions can be used to calculate the strains (ϵ) over the body provided the displacements at the node points are known. Using this fact we can now obtain the corresponding element strains:

$$\epsilon(\mathbf{x}) = \mathbf{B}(\mathbf{x})\mathbf{U} \tag{2}$$

where $\mathbf{B}$ is the strain displacement matrix. The rows of $\mathbf{B}$ are obtained by appropriately differentiating and combining rows of the element interpolation matrix $\mathbf{H}$.

As mentioned earlier, we need to solve the problem of deforming an elastic body to match the set of feature points. This requires solving the *dynamic equilibrium equation*:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{R}, \tag{3}$$

where $\mathbf{R}$ is the load vector whose entries are the spring forces between each feature point and the body surface, and where $\mathbf{M}$ and $\mathbf{K}$ are the element mass and stiffness matrices, respectively.

Both the mass and stiffness matrices are computed directly:

$$\mathbf{M} = \int_V \rho \mathbf{H}^T \mathbf{H} dV \quad \text{and} \quad \mathbf{K} = \int_V \mathbf{B}^T \mathbf{C} \mathbf{B} dV, \tag{4}$$

where ρ is the mass density, and $\mathbf{C}$ is the *material matrix* which expresses the material's particular stress-strain law.

2.2 Mode superposition analysis

This system of equations can be decoupled by posing the equations in a basis defined by the $\mathbf{M}$-orthogonalized eigenvectors of $\mathbf{K}$. These eigenvectors and values are the solution (ϕ_i, ω_i^2) to the following generalized eigenvalue problem:

$$\mathbf{K}\phi_i = \omega_i^2 \mathbf{M}\phi_i. \tag{5}$$

The vector ϕ_i is called the ith *mode shape vector* and ω_i is the corresponding frequency of vibration. Each mode shape vector describes how each node is displaced by the i^{th} vibration mode.

The mode shape vectors ϕ_i are $\mathbf{M}$-orthonormal, this means that

$$\mathbf{\Phi}^T \mathbf{K} \mathbf{\Phi} = \mathbf{\Omega}^2 \quad \text{and} \quad \mathbf{\Phi}^T \mathbf{M} \mathbf{\Phi} = \mathbf{I}. \tag{6}$$

where the ϕ_i are columns in the transform $\mathbf{\Phi}$, and ω_i^2 are the elements of the diagonal matrix $\mathbf{\Omega}^2$.

This generalized coordinate transform $\mathbf{\Phi}$ is then used to transform between nodal point displacements $\mathbf{U}$ and decoupled modal displacements $\tilde{\mathbf{U}}$:

$$\mathbf{U} = \mathbf{\Phi}\tilde{\mathbf{U}} \tag{7}$$

We can now rewrite Equation 3 in terms of these generalized or modal displacements, obtaining a decoupled system of equations:

$$\ddot{\tilde{\mathbf{U}}} + \mathbf{\Omega}^2 \tilde{\mathbf{U}} = \mathbf{\Phi}^T \mathbf{R}, \tag{8}$$

allowing for closed-form solution to the equilibrium problem [5]. Given this equilibrium solution in the two images, point correspondences can be obtained directly.

By discarding high frequency modes the amount of computation required can be minimized without significantly altering correspondence accuracy. Moreover, such a set of modal amplitudes provide a robust, canonical description of shape in terms of deformations applied to the original elastic body. This allows them to be used directly for object recognition [5].

2.3 A New formulation

Perhaps the major limitation of previous methods is that the procedure of attaching virtual springs between data points and the surface of the deformable object implicitly imposes a standard parameterization on the data. We would like to avoid this as much as is possible, by letting the data determine the parameterization in a natural manner.

To accomplish this we use the data itself to define the deformable object, by building stiffness and mass matrices that use the positions of image feature points as the finite element nodes. In [6] we introduced a new finite element formulation which uses Gaussian basis functions as Galerkin interpolants; these interpolants are used to obtain generalized FEM mass and stiffness matrices. Because of space limitations, we will forego a review of this mathematical formulation; readers are referred to [6].We will, however, review our correspondence technique.

3 Determining correspondences

To determine correspondences, we first compute mass and stiffness matrices for both sets of image features. The resulting matrices are then decomposed into eigenvectors ϕ_i and eigenvalues ω_i as described in [5; 6].The resulting eigenvectors are ordered by increasing eigenvalue, and form the columns of the modal matrix

$$\mathbf{\Phi} = \left[\phi_1 \mid \ldots \mid \phi_{2m}\right] = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_m^T \\ \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T \end{bmatrix}, \tag{9}$$

where m is the number of nodes used to build the finite element model. The *mode shape vector* ϕ_i describes the modal displacement (u_i, v_i) at each feature point due to the i^{th} mode, while the *generalized feature vectors* $\mathbf{u}_j$ and $\mathbf{v}_j$ describe the j^{th} feature's displacement in the modal coordinate system.

Correspondences are computed by comparing mode shape vectors for the two sets of features; we will characterize each nodal point by its relative participation in several vibration modes. Before actually describing how this matching is performed, it is important to consider which and how many of these vibration modes should be incorporated into our feature comparisons.

3.1 Modal truncation

For various reasons, we must select a subset of mode shape vectors (column vectors ϕ_i) before computing correspondences. The most obvious reason for this is that the number of eigenvectors and eigenvalues computed for the source and target images will probably not be the same. This is because the number of feature points in each image will almost always differ. To make the dimensions of the two generalized feature spaces the same, we will need to truncate the number of columns at a given dimensionality.

Typically, we retain only the lowest-frequency 25% of the columns of each mode matrix, in part because the higher-frequency modes are the ones most sensitive to noise. Another reason for discarding higher-frequency modes is to make our shape comparisons less sensitive to local shape variations.

We will also want to discard columns associated with the rigid-body modes. Recall that that the columns of the modal matrix are ordered in terms of increasing eigenvalue. For a two-dimensional problem, the first three modes will have near-zero eigenvalues, and will represent two translations and a rotation. These first three columns of each modal matrix are therefore discarded to make the correspondence computation invariant to differences in rotation and translation.

In summary, this truncation breaks the generalized eigenspace into three groups of feature vectors:

$$\begin{aligned} \mathbf{\Phi}_1 &= [\underbrace{\phi_{1,1} \mid \phi_{1,2} \mid \phi_{1,3}}_{\text{rigid body modes}} \mid \underbrace{\phi_{1,4} \mid \ldots \mid \phi_{1,p}}_{\text{intermediate modes}} \mid \underbrace{\ldots \mid \phi_{1,2m}}_{\text{high-order modes}}] \\ \mathbf{\Phi}_2 &= [\overbrace{\phi_{2,1} \mid \phi_{2,2} \mid \phi_{2,3}} \mid \overbrace{\phi_{2,4} \mid \ldots \mid \phi_{2,p}} \mid \overbrace{\ldots \mid \phi_{2,2n}}] \end{aligned} \tag{10}$$

where m and n are the number of features in each image. We keep only those columns which represent the intermediate vibration modes; thus, the truncated generalized feature space will be of dimension $2(p-3)$.

3.2 Computing correspondence affinities

We now have a set of mode-truncated feature vectors for each image. Using a modified version of an algorithm described by Shapiro and Brady [7], we now compute what are referred to as the affinities z_{ij} between the two sets of generalized feature vectors. These are stored in an *affinity matrix* $\mathbf{Z}$, where:

$$\mathbf{z}_{ij} = \| \bar{\mathbf{u}}_{1,i} - \bar{\mathbf{u}}_{2,j} \|^2 + \| \bar{\mathbf{v}}_{1,i} - \bar{\mathbf{v}}_{2,j} \|^2. \tag{11}$$

where the two vectors $\bar{\mathbf{u}}_i$ and $\bar{\mathbf{v}}_i$ store the displacement signature for the i^{th} node point, in truncated mode space.

The affinity measure for the i^{th} and j^{th} points, z_{ij}, will be zero for a perfect match and will increase as the match worsens. Using these affinity measures, we can easily identify which features correspond to each other in the two images by looking for the minimum entry in each column or row of $\mathbf{Z}$.

To obtain accurate correspondences, the Shapiro and Brady method requires three simple, but important, modifications. First, only the generalized features which match with the greatest certainty are used to determine the deformation; the remainder of the correspondences are determined by the deformation itself as in our previous method. By discarding affinities greater than a certain threshold, we allow for tokens which have no strong match. Second, as described earlier, only the low-order twenty-five percent of the eigenvectors are employed, as the higher-order modes are known to be noise-sensitive and thus unstable. And finally, because of the reduced basis matching, similarity of the generalized features is required in both directions, instead of one direction only. In other words, a match between the i^{th} feature in $\mathbf{I}_1$ and the j^{th} feature in $\mathbf{I}_2$ can only be valid if z_{ij} is the minimum value for its row, and z_{ji} the minimum for its column.

3.3 Coping with large rotations

As described so far, our affinity matrix computation method works best when there is little difference in the orientation between images. This is due to the fact that the modal displacements are described as vectors in image space. When the aligning rotation for two sets of features is potentially large, the affinity calculation can be made rotation invariant by transforming the mode shape vectors into a polar coordinate system.

Once each mode shape vector has been transformed into this polar coordinate system, we can compute feature affinities as was described in the previous section. In our preliminary experiments, however, we have found that it is often more effective to compute affinities using either just the r components or just the θ components. In general, the r components are scaled uniformly based on the ratio between the object's overall scale versus the Gaussian basis function radius σ. The θ components, on the other hand, are immune to differences in scale, and therefore a distance metric based on θ offers the advantage of scale invariance.

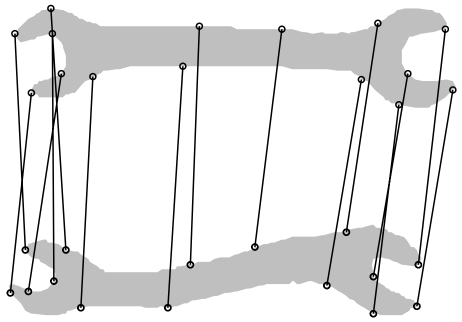


Figure 2: Two wrenches, one straight and one bent, together with the obtained correspondence. The 28 low-order modes were computed for each tree and then correspondences were determined using the algorithm described in the text.

3.4 Correspondence experiments

In this section, we will demonstrate our method's performance on real imagery. In each example the feature points are treated independently; no connectivity or distinctiveness information was employed. Thus the input to the algorithm is a cloud of feature points, not a contour or 2-D form. The mass and stiffness matrices were then computed, and the $\mathbf{M}$-orthonormalized eigenvectors determined. Finally, correspondences were obtained as described above.

Figure 2(a) shows two wrenches: one prototypical straight-handled wrench, and the other a bent wrench. The first 28 modes were computed for both wrenches, and were compared to obtain the correspondences shown. The fact that the two wrenches have similar low-order symmetries (eigenvectors) allows us to recognize that two shapes are closely related, and to easily establish the point correspondences. Roughly 100 silhouette points were matched from each wrench. For clarity, only a few of the computed correspondences are shown in the figure.

In Figure 3, we use outlines of three different types of airplanes as seen from a variety of different viewpoints. In each case, the modal descriptions generated are quite similar, and as a consequence a very good correspondence is obtained. Again, for clarity, only two percent of the correspondences are shown.

4 Alignment, comparison and description

An important benefit of our technique is that the modal vibrations computed for the correspondence algorithm can also be used to describe the rigid and non-rigid deformation needed to align one object with another. Once this *modal description* has been computed, we can compare shapes simply by looking at their mode amplitudes or — since the underlying model is a physical one — we can compute and compare the amount of deformation energy needed to align an object, and use this as a similarity measure. If the modal displacements or strain energy required to align two feature sets is relatively small, then the objects are very similar.

Recall that for a two-dimensional problem, the first three modes will be the rigid body modes of translation and rotation, and the rest are nonrigid modes, and that the nonrigid modes are ordered by increasing frequency of vibration. In general, low-frequency modes describe global deformations, while higher-frequency modes describe more localized shape deformations. This global-to-local ordering of shape deformation allows us to select which types of deformations are to be compared.

For instance, it may be desirable to make object comparisons rotation, position, and/or scale independent. To do this, we ignore displacements in the low-order or rigid body modes, thereby disregarding differences in position, orientation, and scale. In addition, we can make our comparisons robust to noise and local shape variations by discarding higher-order modes. As will be seen later, this modal selection technique is also useful for its compactness, since we can describe deviation from a prototype in terms of relatively few modes.

But before we can actually compare two sets of features, we first need to recover the modal deformations which deform the matched points on one object to their corresponding positions on a prototype object. A number of different methods for recovering the modal deformation parameters are described in the next section.

4.1 Recovering deformations

We want to determine the modal displacements which align the set of features in first image $\mathbf{I}_1$ with corresponding features in second image $\mathbf{I}_2$. Given that $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ have been computed, and that correspondences have been established, then we can solve for the aligning modal displacements directly. This is done by noting that the nodal displacements $\mathbf{U}$ which take features in $\mathbf{I}_1$ to $\mathbf{I}_2$ can be written:

$$\mathbf{u}_i = \mathbf{x}_{1,i} - \mathbf{x}_{2,i}. \tag{12}$$

where corresponding nodes are stored in vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.

Recalling that $\mathbf{U} = \mathbf{\Phi}\tilde{\mathbf{U}}$, and using the identity of Equation 6, we find:

$$\tilde{\mathbf{U}} = \mathbf{\Phi}^T \mathbf{M}\mathbf{U}. \tag{13}$$

Normally we do not have one-to-one correspondences between the features. In cases where the recovery is overconstrained, we can solve via least squares. In cases where the recovery is underconstrained, we would like unmatched nodes move in a manner consistent with the material properties and the forces at the matched nodes. This type of solution is obtained by setting the loads $\mathbf{r}_i$ pulling on unmatched nodes to zero. We then solve for the unknown displacements by plugging into the equilibrium equation:

$$\mathbf{K}\mathbf{U} = \mathbf{R}, \tag{14}$$

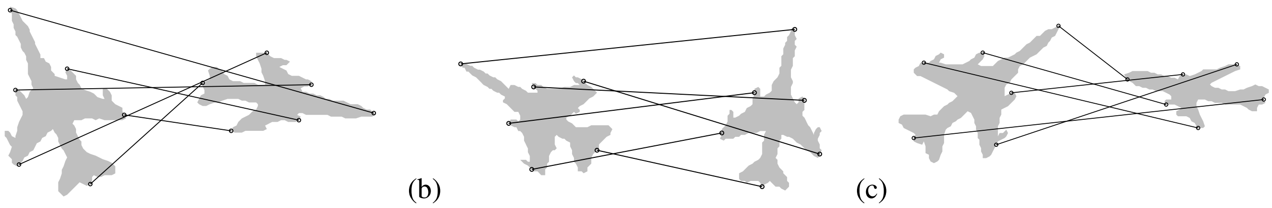
where we have as many knowns as unknowns.

Alternatively, we can solve for the modal displacements by using a mode-truncated version of Equation 13. Assume that we have found correspondences for p of the m nodes. We reduce the degrees of freedom by discarding $m - p$ of the high-frequency eigenvectors (columns ϕ_i); this yields a mode-truncated transformation matrix $\bar{\mathbf{\Phi}}$. We then discard the rows of $\bar{\mathbf{\Phi}}$ which are associated with unmatched nodes. The resulting system of equations has $2p$ equations and $2p$ unknowns, so we can solve for the modal displacements directly. We again obtain a solution in closed-form, but we have assumed that the modal displacements $\tilde{\mathbf{u}}_i = 0$, for $i > p$. By requiring that strain-energy be minimized, we can use an iterative technique for obtaining non-zero $\tilde{\mathbf{u}}_i$, for $i > p$.

When solving in modal space, the minimum strain energy can be measured directly in terms of modal displacements, *i.e.:*

$$E_I = \frac{1}{2}\tilde{\mathbf{U}}^T \mathbf{\Omega}^2 \tilde{\mathbf{U}}. \tag{15}$$

This strain energy equation enforces a penalty which is proportional to the squared vibration frequency associated with each mode. Since rigid body modes ideally introduce no strain, it is logical that their $\omega_i \approx 0$.



(a) (b) (c)

Figure 3: Correspondence obtained for outlines of different types of airplanes. The first example shows the correspondences found for different (rotated in 3D) views of the same fighter plane. The others show matches between increasingly different airplanes. Roughly 150 silhouette points were matched from each plane. Because of the large number of data points, only critical correspondences are shown.

4.2 Comparing objects

Once the mode amplitudes have been recovered, we can compute the strain energy incurred by these deformations by plugging into Equation 15. This strain energy can then be used as a similarity metric. As will be seen in the examples, we may also want to compare the strain in a subset of modes only, or the strain for each mode separately. The strain associated with the i^{th} mode is simply:

$$E_{mode_i} = \frac{1}{2}\tilde{u}_i^2 {\omega_i}^2. \tag{16}$$

Since each mode's strain energy is scaled by its frequency of vibration, there is an inherent penalty for deformations which occur in the higher-frequency modes. In our experiments, we have used strain energy for most of our object comparisons, since it has a convenient physical meaning; however, we suspect that (in general) it will be necessary to weigh higher-frequency modes less heavily, since these modes typically only describe local shape variations.

4.3 Alignment and description

Figure 4 demonstrates how we can align a prototype shape with other shapes, and how to use this computed strain energy as a similarity metric. As input, we are given the correspondences computed for the various airplane silhouettes shown in Figure 3. Our task is to align and describe the three different target airplanes (shown in gray) in terms of modal deformations of a prototype airplane (shown in black). In each case, there were approximately 150 contour points used, and correspondences were computed using the first 36 vibration modes. On the order of 50 strongest corresponding features were used as input to the strain-minimizing version of the Equation 13. The modal strain energy was computed using Equation 15.

The graphs in Figure 4 show the values for the 36 recovered modal amplitudes needed to align or warp the prototype airplane with each of the target airplanes. These mode amplitudes are essentially a recipe for how to build each of the three target airplanes in terms of deformations from the prototype.

Figure 4(a) shows an airplane that is similar to the prototype, and which is viewed from a viewpoint that results in a similar image geometry. As a consequence, the two planes can be accurately aligned with little deformation, as indicated by the graph of mode amplitudes required to warp the prototype to the target shape.

Figure 4(b) depicts an airplane which is from the same class of airplanes as the prototype, but viewed from a very different angle. In this case, the graph of deformation mode amplitudes shows a sizable strain in the first few modes. This makes sense, since generally the first six to nine deformation modes account for affine-like deformations which are similar to the deformations produced by changes in viewpoint.

The final example, Figure 4(c), is very different from the prototype airplane, and is viewed from a different viewpoint. In this case, the recovered mode deformations are large in both the low and higher-frequency modes.

This figure illustrates how the distribution of deformation energy in the various modes can be used judge the similarity of different shapes, and to determine if differences are likely due primarily to changes in viewpoint. Figure 4(a) shows that similar shapes can be aligned with little deformation; (b) shows that viewpoint changes produce mostly low-frequency deformations, and (c) shows that to align different shapes generally requires deformations of both low and high frequency.

4.4 Determining relationships between objects

By looking more closely at the mode strains, we can pin-point which modes are predominant in describing an object. Figure 5 shows what we mean by this. As before, we can describe one object's silhouette features in terms of deformations from a prototype. In this case, we want to compare different hand tools. The prototype is a wrench, and the two target objects are a bent wrench and hammer. Silhouettes were extracted from the images, and thinned down to between 60 and 120 points per contour. Using the strongest matched contour points, we then recovered the first 28 modal deformations which warp the prototype onto the other tools. The strain energy attributed to each modal deformation is shown in the graph at the bottom of the figure. As can be seen from the graph, the energy needed to align the prototype with a similar object (the bent wrench) was mostly isolated in two modes: modes 6 and 8. In contrast, the strain energy needed to align the wrench with the hammer is much greater and spread across the graph.

Figure 6 shows the result of aligning the prototype with the two other tools using only the two most dominant modes. The top row shows alignment with the bent wrench using just the sixth mode (a shear), and then just the eighth mode (a simple bend). Taken together, these two modes do a very good job of describing the deformation needed to align the two wrenches. In contrast, aligning the wrench with the hammer (bottom row of Figure 6) cannot be described simply in terms of a few deformations of the wrench.

By observing that there is a simple physical deformation which aligns the prototype wrench and the bent wrench, we can conclude that they are probably closely related in category and functionality. In contrast, the fact that there is no simple physical relationship between the hammer and the wrench indicates that they are likely to be different types of object, and may have quite different functionality.

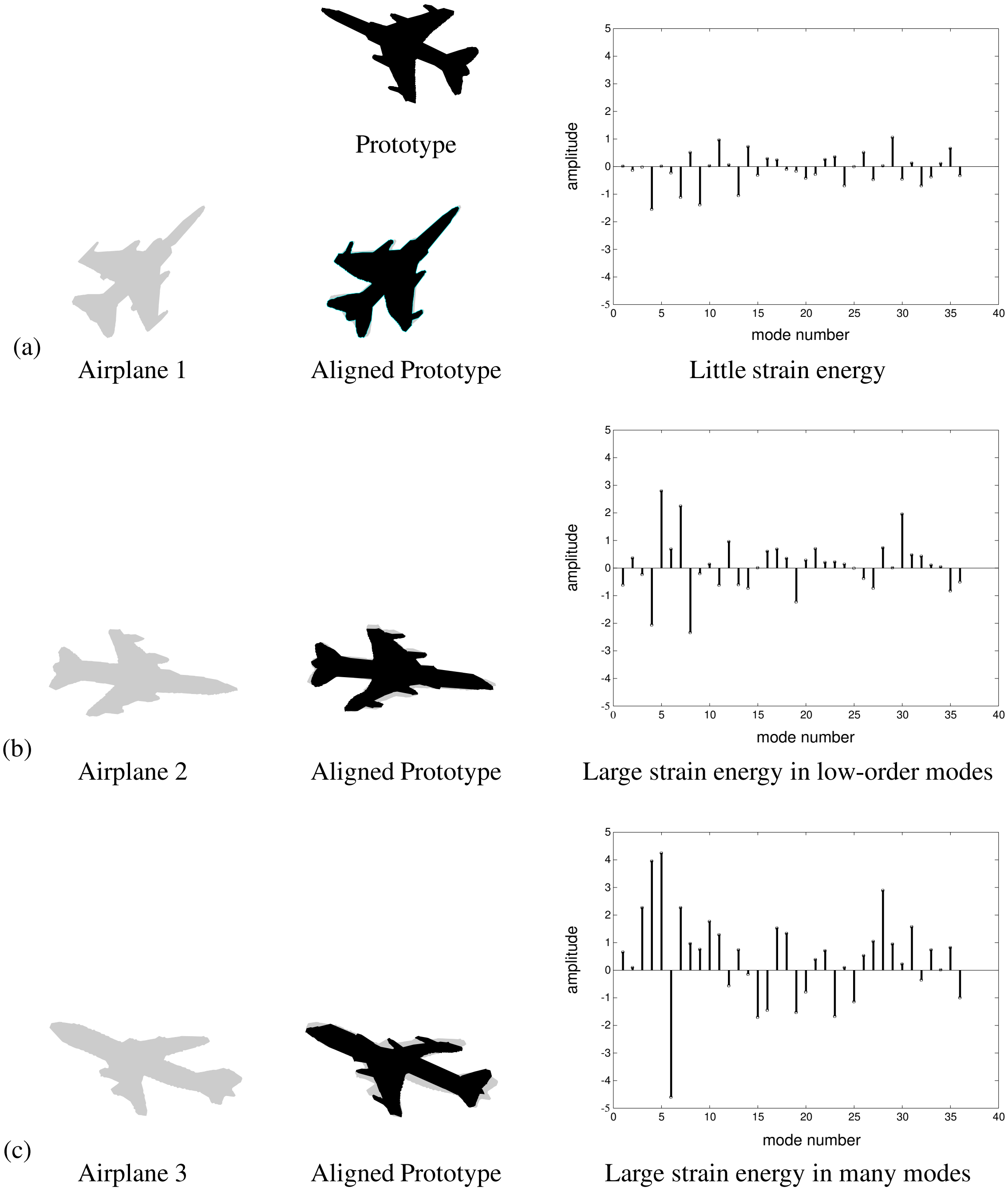


Figure 4: Describing planes in terms of a prototype. The graphs show the 36 mode amplitudes used to align the prototype with each target shape. (a) shows that similar shapes can be aligned with little deformation; (b) shows that viewpoint changes produce mostly low-frequency deformations, and (c) shows that to align different shapes requires both low and high frequency deformations.

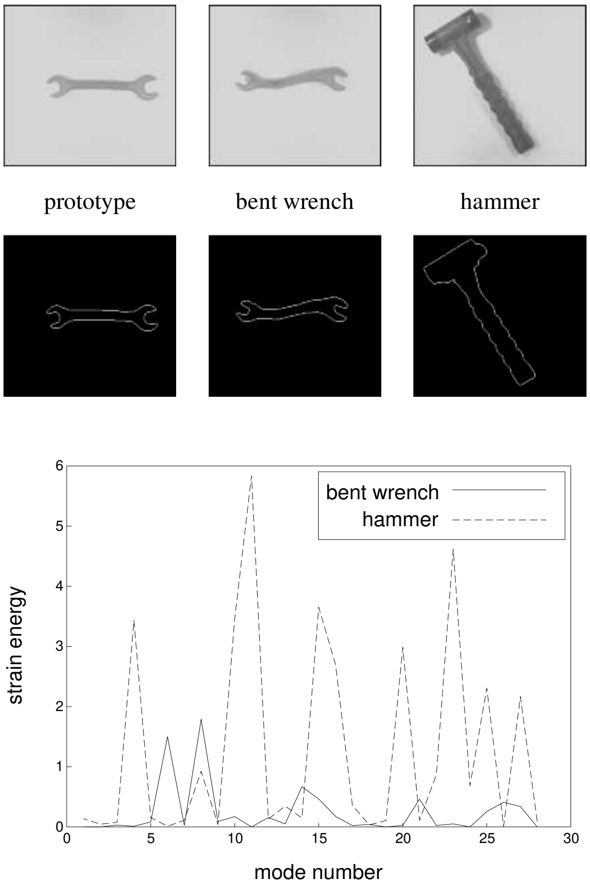


Figure 5: Describing a bent wrench and a hammer in terms of modal deformations from a prototype wrench. Silhouettes were extracted from the images, and then the strongest corresponding contour points were found. Using these matched contour points, the first 28 modal deformations which warp the prototype's contour points onto the other tools were then recovered and the resulting strain energy computed. A graph of the *modal strain* attributed to each modal deformation is shown at the bottom of the figure.

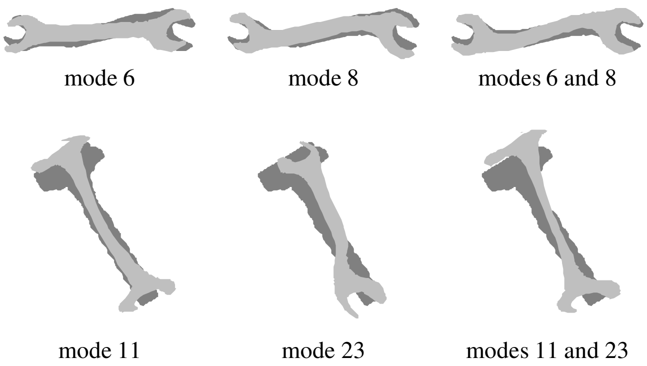


Figure 6: How the top two highest-strain modal deformations contribute to the alignment of a prototype wrench to the bent wrench and a hammer of Figure 5.

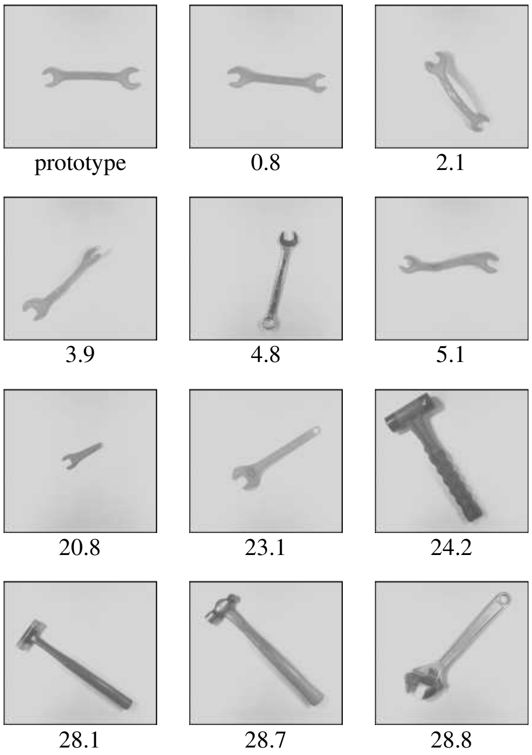


Figure 7: Using modal strain energy to compare a prototype wrench with different hand tools. Silhouettes were first extracted from each tool image, and then the strongest corresponding contour points were found. The first 22 mode amplitudes were recovered and used to warp the prototype onto the other tools. The resulting modal strain energy is shown below each image. As can be seen, strain energy provides an good measure for similarity.

4.5 Recognition of objects and categories

In the next example (Figures 7 and 8) we will use modal strain energy to compare three different prototype tools: a wrench, hammer, and crescent wrench. As before, silhouettes were first extracted and thinned from each tool image, and then the strongest corresponding contour points were found. Mode amplitudes for the first 22 modes were recovered and used to warp each prototype onto the other tools. The modal strain energy which results from deforming the prototype to each tool is shown below each image.

Figure 7 depicts the use of modal strain energy in comparing a prototype wrench with eleven other hand tools. As this figure shows, the shapes most similar to the wrench prototype are those other two-ended wrenches with approximately straight handles. Next most similar are closed-ended and bent wrenches, and most dissimilar are hammers and single-ended wrenches. Note that the matching is orientation and scale invariant (modulo limits imposed by pixel resolution).

Figure 8 continues this example using as prototypes the hammer and a single-ended wrench. Again, the modal strain energy which results from deforming the prototype to each tool is shown below each image.

When the hammer prototype is used, the most similar shapes found are three other images of the same hammer, taken with different viewpoints and illumination. The next most similar shapes are a variety of other hammers. The least similar shapes are a set of wrenches.

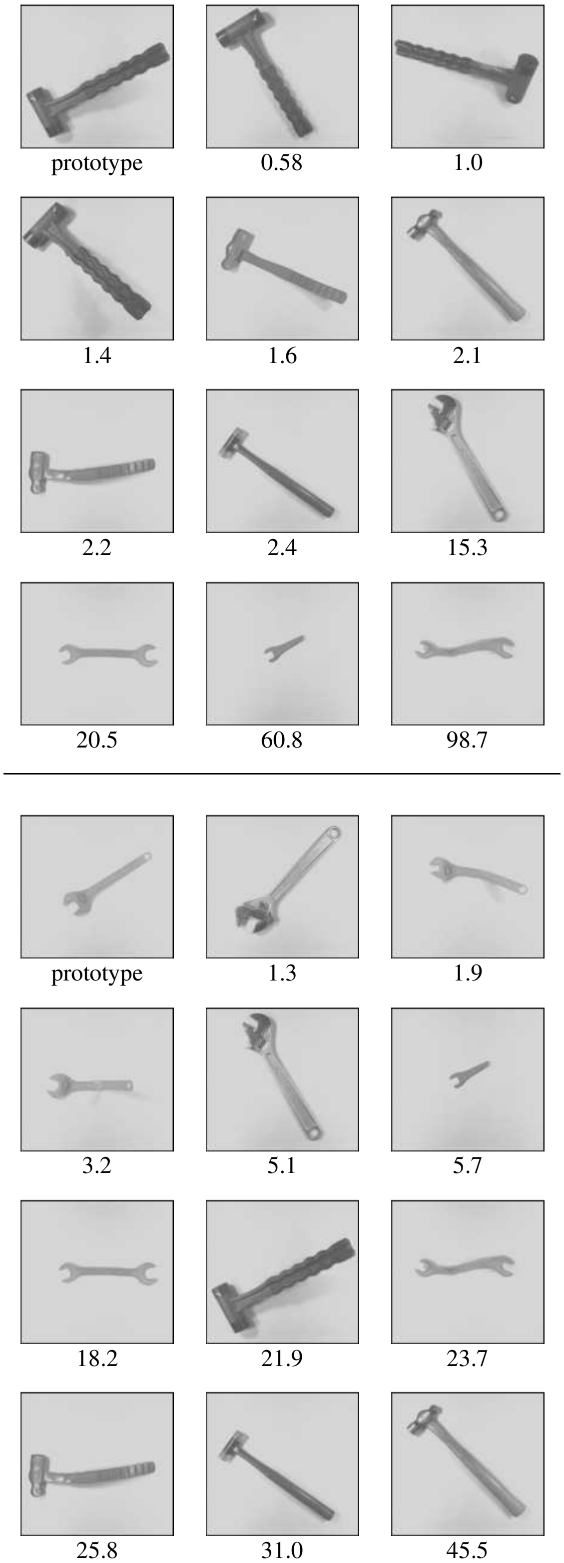


Figure 8: Using modal strain energy to compare a crescent wrench with different hand tools, and a prototype hammer with different hand tools. The modal strain energy which results from deforming the prototype to each tool is shown below each image.

For the single-ended wrench prototype, the most similar shapes are a series of single-ended wrenches. The next most similar is a straight-handled double-ended wrench, and the least similar are a series of hammers and a bent, double-ended wrench.

The fact that the similarity measure produced by the system corresponds to functionally-similar shapes is important. It allows us to recognize the most similar wrench or hammer from among a group of tools, even if there is no tool that is an exact match. Moreover, if for some reason the most-similar tool can't be used, we can then find the next-most-similar tool, and the next, and so on. We can find (in order of similarity) all the tools that are likely to be from the same *category*.

5 Conclusion

The transformation to modal space not only allows for automatically establishing correspondence between clouds of feature points; the same modes (and the underlying FEM model) can be used to describe the deformations which take the features from one position to the other. The amount of deformation required to align the two feature clouds can be used for shape comparison and description, and to warp the original images for alignment and sensor fusion.

The descriptions computed are canonical, and vary smoothly even for very large deformations. This allows them to be used directly for object recognition as illustrated by the airplane and hand-tool examples in the previous section. Finally, because the deformation comparisons are physically-based, we can determine whether or not two shapes are related by a simple physical deformation. This has allowed us to identify shapes that appear to be members of the same category.

References

[1] T. Cootes, D. Cooper, C. Taylor, and J. Graham. Trainable method of parametric shape description. *Image and Vision Computing*, 10(5):289–294, June 1992.

[2] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1:321–331, 1987.

[3] P. Lipson, A. Yuille, D. O'Keefe, J. Cavanaugh, J. Taafe, and D. Rosenthal. Deformable templates for feature extraction from medical images. In *Proc. ECCV*, pages 413–417, 1990.

[4] A. Pentland. Automatic extraction of deformable part models. *IJCV*, pages 107–126, 1990.

[5] A. Pentland and S. Sclaroff. Closed-form solutions for physically-based shape modeling and recognition. *IEEE TPAMI*, 13(7), July 1991.

[6] S. Sclaroff and A. Pentland. A modal framework for correspondence and recognition. In *Proc. ICCV*, May 1993.

[7] L. Shapiro and J. Brady. Feature-based correspondence: an eigenvector approach. *Image and Vision Computing*, 10(5):283–288, June 1992.

[8] L. Staib and J. Duncan. Parametrically deformable contour models. In *Proc. CVPR*, pages 98–103, 1989.

[9] D. Terzopoulos and D. Metaxas. Dynamic 3d models with local and global deformations: Deformable superquadrics. *IEEE TPAMI*, 13(7):703–714, July 1991.

[10] D. Terzopoulos, A. Witkin, and M. Kass. Constraints on deformable models: Recovering 3d shape and nonrigid motion. *Artificial Intelligence*, 36:91–123, 1988.