Search by Shape Examples: Modeling Nonrigid Deformation

Stan Sclaroff *
Computer Science Dept.
Boston University
111 Cummington St., Boston MA 02215

Alex P. Pentland
Perceptual Computing Section
The MIT Media Laboratory
20 Ames St., Cambridge MA 02139

Abstract

We describe our work on shape-based image database search using the technique of modal matching. Modal matching employs a deformable shape decomposition that allows users to select example objects and have the computer efficiently sort the set of objects based on the similarity of their shape. Shapes are compared in terms of the types of nonrigid deformations (differences) that relate them. The modal decomposition provides deformation "control knobs" for flexible matching and thus allows for selecting weighted subsets of shape parameters that are deemed significant for a particular category or context. We demonstrate the utility of this approach for shape comparison in 2-D image databases; however, the general formulation is applicable to signals of any dimensionality.

1 Introduction

Automated image database search requires that human users be able to communicate their goals to the computer. The problem is how to convey this information: while humans perceive shape and structure in an image, to the computer image data is merely an array of bits. One way to bridge this communications gap is to allow users to select example images of what to "look" for. However, even when given examples, recognizing and then interpreting what is important in an image remains a critical computing challange.

In the last few years researchers have made some progress toward automatic shape indexing for image databases. The general approach has been to calculate some approximately invariant statistic like shape moments, and use these to stratify the image database [1, 2, 3, 4, 5, 9].

The problem with this general approach is that it discards significant perceptual and semantic information. Rather than discarding useful similarity information by employing only invariants, we believe that one should use a decomposition that preserves as much semantically meaningful and perceptually important information as is possible, while still providing an efficient encoding of the original signal [6]. We argue the image database problem requires having an arsenal of such decompositions, each specially trained for describing a particular type of object or context (*e.g.*, the Karhunen-Loève transform for faces [12], or the Wold decomposition for textures [8]).

In this paper, we describe *modal matching*, an information-preserving shape decomposition for matching, describing, and comparing shapes despite sensor variations and deformations. Modal matching employs a shape decomposition that allows users to select examples, and then has the computer efficiently compare shapes in terms of the types of nonrigid deformations (differences) that relate them. Modal matching utilizes the eigenvectors of the finite element stiffness matrix, a positive definite matrix that describes the connectedness between features. These eigenvectors provide a new, generalized coordinate system for describing the location of feature points.

*This work was done while the author was at the Media Lab.

Since the underlying representation is based on the finite element method, optimal estimates of object motion and shape can be made, and physical predictions and simulations can be computed directly from recovered models. We will demonstrate the utility of this approach for comparing shapes in 2-D image databases of animals and hand tools, based on feature point and silhouette data.

2 Review: The Modal Representation

A shape's modal representation is based on the eigenvectors of its physical model. The mathematical formulation of this physical model is based on the finite element method (FEM), the standard engineering technique for simulating the dynamic behavior of an object. In the FEM, interpolation functions are developed that allow continuous material properties, such as mass and stiffness, to be integrated across the region of interest. Solution to the problem of deforming an elastic body to match the set of feature points requires solving the *dynamic equilibrium equation*:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{R}, \tag{1}$$

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. For an in-depth description of this formulation, readers are directed to [7, 10, 11].

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:

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

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

$$\mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = \mathbf{\Omega}^2 \quad \text{and} \quad \mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi} = \mathbf{I}. \tag{3}$$

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 $\tilde{\mathbf{U}}$, $\mathbf{U} = \mathbf{\Phi}\tilde{\mathbf{U}}$. We can now rewrite Eq. 1 in terms of these generalized or modal displacements, obtaining a decoupled system of equations:

$$\ddot{\tilde{\mathbf{U}}} + \mathbf{\Omega}^2\tilde{\mathbf{U}} = \mathbf{\Phi}^T\mathbf{R}, \tag{4}$$

allowing for closed-form solution to the equilibrium problem [7].

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 [7].

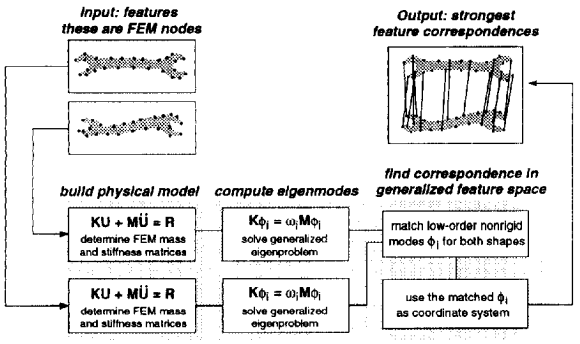

Figure 1: System diagram.

3 Modal Matching

Imagine that we are given two sets of image feature points, and that our goal is to determine if they are from two similar objects. The most common approach to this problem is to try to find distinctive local features that can be matched reliably; this fails because there is insufficient local information, and because viewpoint and deformation changes can radically alter local feature appearance.

An alternate approach is to first determine a body-centered coordinate frame for each object, and then attempt to match up the feature points. Once we have the points described in intrinsic or *body-centered* coordinates rather than Cartesian coordinates, it is easy to match up the bottom-right, top-left, etc. points between the two objects. Modal matching [10] provides provides such a body-centered coordinate system.

A flow-chart of our method is shown in Fig. 1. For each image we start with feature point locations $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_m]$ and use these as nodes in building a finite element model of the shape. We can think of this as constructing a model of the shape by covering each feature point with a Gaussian blob of rubbery material; if we have segmentation information, then we can fill in interior areas and trim away material that extends outside of the shape.

We then compute the eigenmodes (eigenvectors) ϕ_i of the finite element model. The eigenmodes provide an orthogonal frequency-ordered description of the shape and its natural deformations. They are sometimes referred to as *mode shape vectors* since they describe how each mode deforms the shape by displacing the original feature locations.

The first three eigenmodes are the rigid body modes of translation and rotation, and the rest are nonrigid modes. 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 will prove very useful for shape matching and comparison.

The eigenmodes 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 eigenmode*. The transform between Cartesian feature locations and modal feature locations is accomplished by using the FEM eigenvectors as a coordinate basis. In our technique, two groups of features are compared in this eigenspace. The important idea here is that the low-order modes computed for two similar objects will be very similar — even in the presence of affine deformation, nonrigid deformation, local shape perturbation, or noise.

Using this property, feature correspondences are found via *modal matching*. The concept of modal matching is demonstrated on the two similar tree shapes in Fig. 2. Correspondences are found by comparing the direction of displacement at each node. The direction of displacement is shown by vectors in figure. For instance, the top points on the two trees in Fig. 2(a, b) have very similar displacements across a number of low-order modes, while the bottom point (Fig. 2(c)) has a very different displacement signature. Good matches have similar displacement signatures, and so the system matches the top points on the two trees.

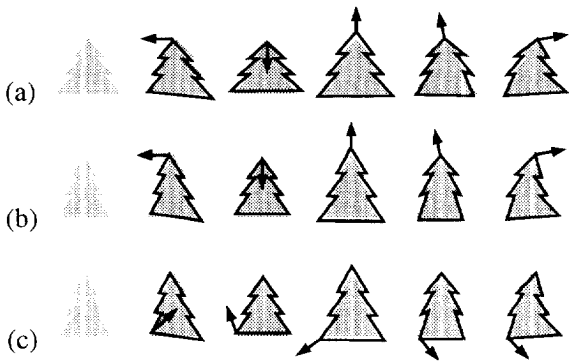

Figure 2: Computing correspondences in modal signature space. Given two similar shapes, correspondences are found by comparing the direction of displacement at each node (shown by vectors in figure). For instance, the top points on the two trees (a, b) have very similar displacement signatures, while the bottom point (shown in c) has a very different displacement signature. Using this property, we can reliably compute correspondence affinities in this modal signature space.

Point correspondences between two shapes can be reliably determined by comparing their trajectories in this modal space. In the implementation described in this paper, points that have the most similar unambiguous coordinates are matched via modal matching, with the remaining correspondences determined by using the physical model as a smoothness constraint. Currently, the algorithm has the limitation that it cannot reliably match largely occluded or partial objects.

4 Modal Descriptions

An important benefit of our technique is that the eigenmodes 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.

4.1 Recovering deformations

Before we can actually compare two sets of features, we first need to recover the modal deformations $\tilde{\mathbf{U}}$ that deform the matched points on one object to their corresponding positions on a prototype object. Given that modal models have been computed for both shapes, and that correspondences have been established, then we can solve for the modal displacements directly. This is done by noting that the nodal displacements $\mathbf{U}$ that align corresponding features on both shapes can be written: $\mathbf{u}_i = \mathbf{x}_{1,i} - \mathbf{x}_{2,i}$, where $\mathbf{x}_{1,i}$ is the i^{th} node on the first shape and $\mathbf{x}_{2,i}$ is its matching node on the second shape.

Normally there is not one-to-one correspondence between the features. In the more typical case where the recovery is underconstrained, we would like unmatched nodes to move in a manner consistent with the material properties and the forces at the matched nodes. This type of solution can be obtained via strain-minimizing least squares.

The strain energy can be measured directly in terms of modal displacements, and enforces a penalty that is proportional to the squared vibration frequency associated with each mode:

$$E_I = \frac{1}{2}\tilde{\mathbf{U}}^T\Omega^2\tilde{\mathbf{U}}. \quad (5)$$

We now formulate a strain-minimizing least squares solution, where we minimize alignment error that includes this modal strain energy term:

$$\tilde{\mathbf{U}} = \left[\mathbf{\Phi}^T\mathbf{W}^2\mathbf{\Phi} + \lambda\Omega^2\right]^{-1}\mathbf{\Phi}^T\mathbf{W}^2\mathbf{U}, \quad (6)$$

where $\mathbf{W}$ is a diagonal matrix whose entries are inversely proportional to the affinity measure for each feature match. This approach exploits the underlying physical model to enforce certain geometric constraints in a least squares solution.

4.2 Comparing objects

Once mode amplitudes have been recovered, we can compute the strain energy incurred by these deformations via Eq. 5. 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: $\frac{1}{2}\tilde{u}_i^2{\omega_i}^2$.

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.

Since each mode's strain energy is scaled by its frequency of vibration, there is an inherent penalty for deformations that 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 are more susceptible to noise.

4.3 Modal shape categories

Using methods similar to those employed by Ullman and Basri [13] we can describe objects as linear combinations of some collection of base models. The difference here is that we have a frequency-ordered description of shape; as a result we can analyze and decompose nonrigid shape deformation (and then synthesize shapes) in a principled way.

Fig. 3(a) shows a *shape space* defined by three prototype models. Using modal matching, correspondences were determined and similarity (strain energy) was computed. Each edge is labelled with its associated strain. Traveling along an edge in this triangle performs a linear blend, using the modal deformations, from one prototype model to another. Thus, each edge of the triangle describes a family of models that can be represented as linear combinations of the two prototypes. Similarly, we can describe an entire family of shapes by moving around inside the triangle defined by three models.

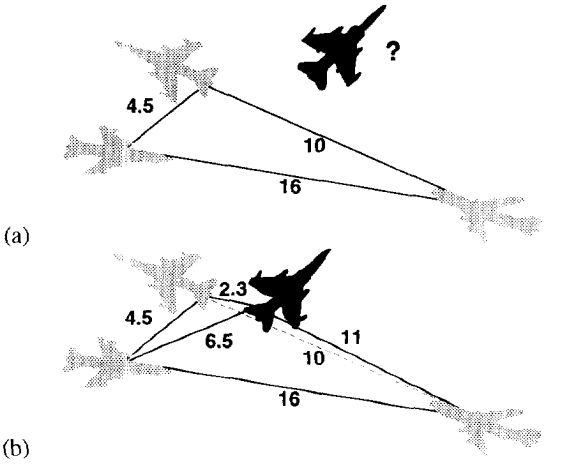

Figure 3: Three gray models define a triangle (a) with edge lengths proportional to the amount of strain needed to align each model. A pyramid (b) results when a fourth model cannot be completely explained by the three known models.

Adding a fourth model to the triangle creates a pyramid, unless the new model can be exactly described as a linear combination of the prototype models. Fig. 3(b) shows how the fourth plane model was synthesized from a combination of the three base models. The three base models cannot completely account for all of the new plane's shape (there are are missing nacelles, for instance). The distance between the new plane and the triangle of base shapes is the similarity between the new plane and the *class* of shapes defined by linear combinations of the prototype models. Using this similarity measure, we can decide whether or not the new shape is a member of the class defined by the prototype models.

5 Examples

5.1 Determining relationships between objects

By looking closely at the mode strains, we can pin-point which modes are predominant in describing an object. Fig. 4 uses this principle to compare different handtools. 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 that 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.

Fig. 5 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 Fig. 5) cannot be described simply in terms of a few deformations of the wrench.

By observing that there is a simple physical deformation that 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

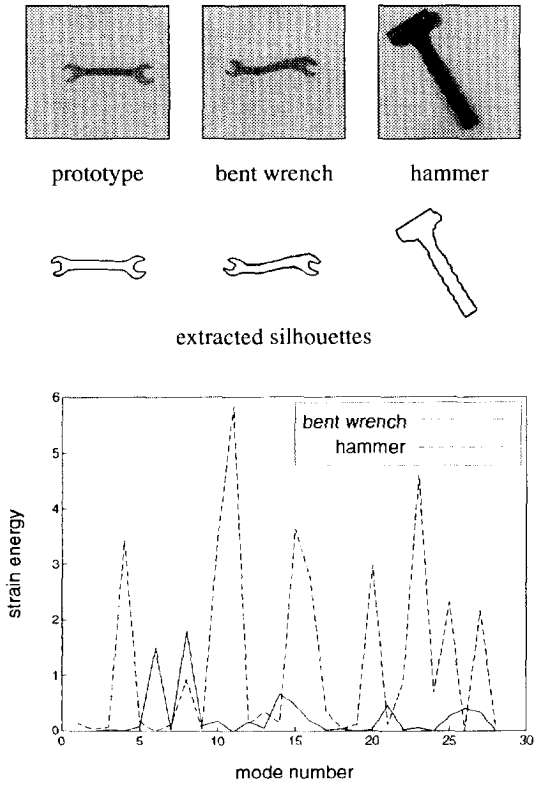

Figure 4: Describing a bent wrench and a hammer in terms of modal deformations from a prototype wrench. Silhouettes were extracted from the images, and then the first 28 modal deformations that warp the prototype's contour points onto the other tools were recovered. A graph of the *modal strain* attributed to each modal deformation is shown at the bottom of the figure.

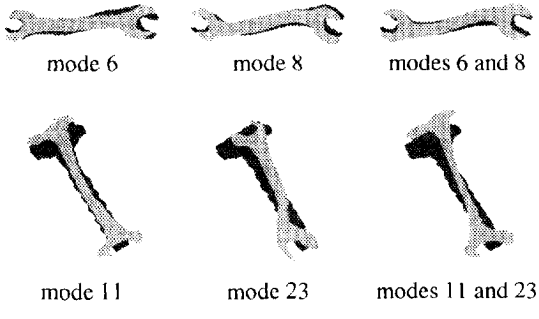

Figure 5: How the top two highest-strain modal deformations contribute to the alignment of a prototype wrench to the bent wrench and a hammer of Figure 4.

between the hammer and the wrench indicates that they are likely to be different types of object.

5.2 Recognition of shape categories

In the next example (Fig. 6) we will use modal strain energy to compare two different prototype tools: a wrench and a hammer. 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 28 modes were recovered and used to warp each prototype onto the other tools. The modal strain energy that results from deforming the prototype to each tool is shown below each image.

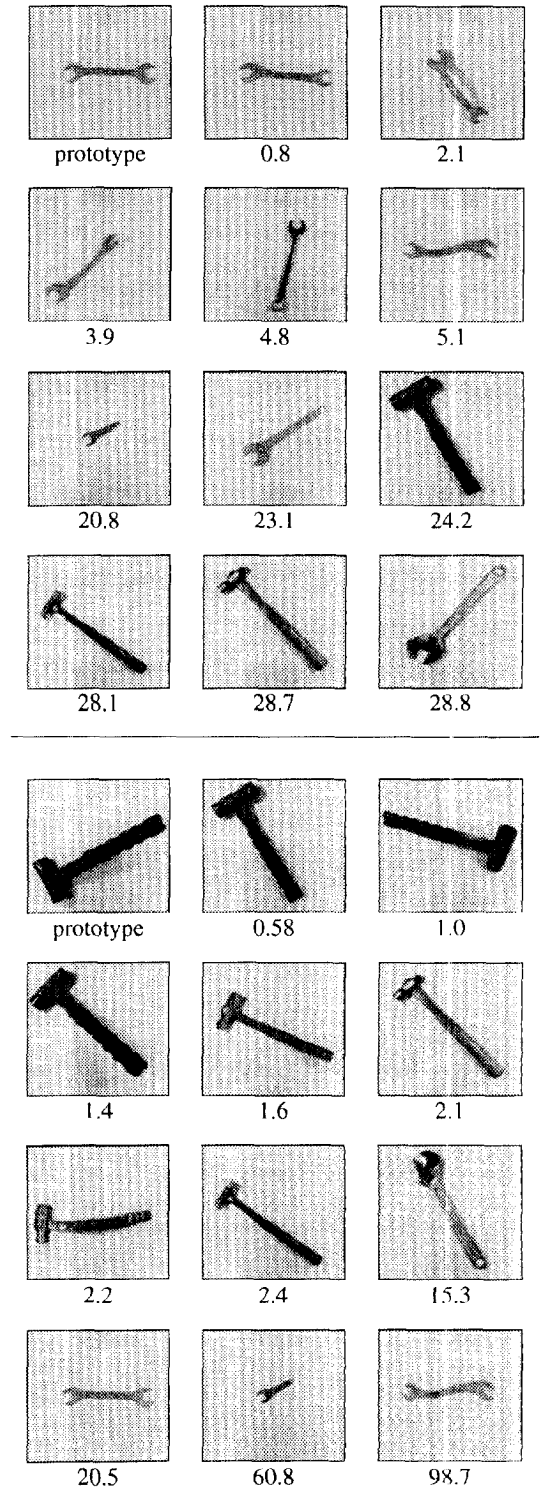

Figure 6: Using modal strain energy to compare prototype wrench with different hand tools, and a protoype hammer with different hand tools. Silhouettes were first extracted from each tool image, and then the strongest corresponding contour points were found. The first 28 mode amplitudes were recovered 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.

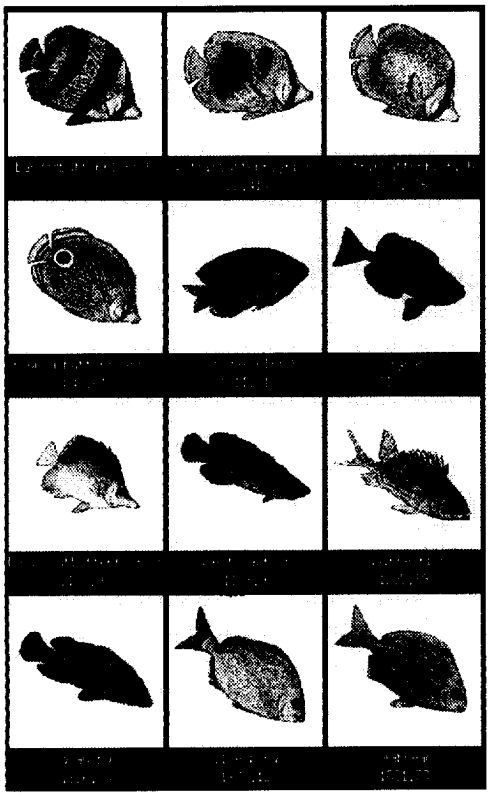

Figure 7: Ordering fish shapes in terms of distances to three prototype fish. Each fish shape in the database was matched and warped to three different prototype fish shapes, and the resulting modal strain energy stored as a three-tuple. Based on these strain coordinates, the system retreived the fish shapes that were closest to the banded butterfly fish shape (other butterfly fish).

As this Fig. 6 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).

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.

The fact that the similarity measure produced by the system 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.

The last example shows our most recent progress towards structuring image databases into categories in terms of a few prototype shapes. As described in Sec. 4.3, a category of shapes can be represented as linear combinations of a small representative collection of base models. Fig. 7 depicts a preliminary result in applying this to a database of tropical fish images.

Each fish shape in the database was matched and warped to three different prototype fish shapes, and the resulting modal strain energy stored as a three-tuple. This tuple maps the shapes into a three-dimensional space where each axis represents one of the prototype shapes. The number shown below each image in the figure is the Euclidean distance in this prototype strain-space. The matches are shown in order, starting with the most similar. Based on these distances, the system retreived the fish shapes that were closest to the banded butterfly fish shape (other butterfly fish).

6 Conclusion

Modal matching employs a shape decomposition that allows users to select examples, and then has the computer efficiently match and compare shapes in terms of an ordered set of orthogonal deformation *modes*. In the modal method, shape information is decomposed into an ordered basis of orthogonal principal components. As a result, the less critical and often noisy high-order components can be discarded in order to obtain overconstrained, canonical descriptions. This allows for the selection of only the most important components to be used for efficient data reduction, real-time recognition, and robust reconstruction. 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] Z. Chen and S. Y. Ho. Computer vision for robust 3D aircraft recognition with fast library search. *Pattern Recognition*, 24(5):375–390, 1991.

[2] M. A. Ireton and C. S. Xydeas. Classification of shape for content retrieval of images in a multimedia database. In *Proc. Int. Conf. on Digital Proc. of Sig. in Com.*, Sep. 1990.

[3] H. V. Jagadish. A retrieval technique for similar shapes. In *Proc. Int. Conf. on Management of Data, ACM SIGMOD 91*, May 1991.

[4] R. Mehrotra and W. I. Grosky. Shape matching utilizing indexed hypotheses generation and testing. *IEEE Trans. on Robotics and Automation*, 5(1):70–77, 1989.

[5] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, and P. Yanker. The QBIC project: Querying images by content using color, texture, and shape. In *Proc. SPIE Conf. on Storage and Retrieval of Image and Video Databases*, 1908, Feb. 1993.

[6] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. In *Proc. SPIE Conf. on Storage and Retrieval of Image and Video Databases II*, 2185, Feb. 1994.

[7] A. Pentland and S. Sclaroff. Closed-form solutions for physically-based shape modeling and recognition. *IEEE PAMI*, 13(7):715–729, Jul. 1991.

[8] R. Picard and F. Liu. A new Wold ordering for image similarity. In *Proc. ICASSP*, Apr. 1994.

[9] B. Scassellati, S. Alexopoulos, and M. Flickner. Retrieving images by 2D shape: comparison of computation methods with human perceptual judgements. In *Proc. SPIE Conf. on Storage and Retrieval of Image and Video Databases II*, Feb. 1994.

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

[11] S. Sclaroff and A. Pentland. Modal matching for correspondence and recognition. *IEEE PAMI*, in press.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cog. Neuroscience*, 3(1):71–86, 1991.

[13] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE PAMI*, 13(10):992–1006, 1991.