SHAPE PRIORS BY KERNEL DENSITY MODELING OF PCA RESIDUAL STRUCTURE

J.P. Lewis
Stanford University
Computer Graphics Lab

Iman Mostafavi
UC San Diego
Computer Science and Engineering

Gina Sosinsky, Maryanne E. Martone, Ruth West
National Center for
Microscopy and Imaging Research

ABSTRACT

Modern image processing techniques increasingly use prior models of the expected distribution of objects. Principal component eigenmodels are often selected for shape prior modeling, but are limited in capturing only the second order moment statistics. On the other hand, kernel densities can in concept reproduce arbitrary statistics, but are problematic for high dimensional data such as shapes. An evident approach is to combine these methods, using PCA to reduce the problem dimensionality, followed by kernel density modeling of the PCA coefficients. In this paper we show that useful algorithmic and editing operations can be formulated in term of this simple approach. The operations are illustrated in the context of point distribution shape models. Particular points can be rapidly evaluated as being plausible or outliers, and a plausible shape can be completed given limited operator input in a manually guided procedure. This "PCA+KD" approach is conceptually simple, scalable (becoming increasingly accurate with additional training data), provides improved modeling power, and supports useful algorithmic queries.

Index Terms— Shape analysis, priors, segmentation.

1. INTRODUCTION

Prior models help an image processing system distinguish between signal and noise, and serve to select among the solutions in underconstrained and inverse problems. Priors are widely used, whether explicitly formulated as in Bayesean approaches, or implicit in an algorithm. Because of their importance and wide use, the subject of priors merits independent study.

The major challenge in developing a shape or image prior model is that the space is high dimensional. The set of "valid" shapes is difficult to characterize, as it is a small and nonlinear subspace of the overall space. Early priors considered only local (typically smoothness) properties rather than modeling the complete shape, thereby reducing the dimensionality of the problem. More recently, global principal component analysis (PCA) methods have targeted full shapes but model only a subset of the possible statistics, specifically the second-order moments. Extending shape models beyond these Gaussian statistics is a subject of recent and current research.

In this paper we consider the approach of employing kernel density (KD) to model the residual non-Gaussian distribution of PCA coefficients. The modeling power is thus extended beyond second-order statistics, while the "curse of dimensionality" that affects kernel density modeling is partially addressed through the dimensionality reduction afforded by PCA.

This PCA+KD approach to priors has the advantages of being conceptually simple, scalable (becoming increasingly accurate as the available training data grows), and having potentially increased modeling power.

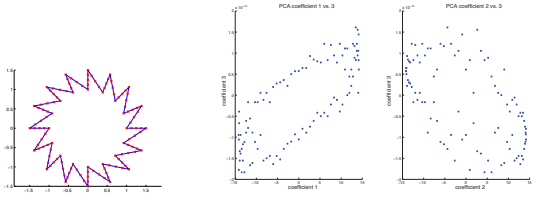

Figure 1. Directional characteristics of this simple shape class (left) result in clear non-Gaussian structure in the PCA coefficients (right). Plots of coefficients 1 vs. 3 and 2 vs. 3 for a set of rotated versions of this shape are shown.

We apply our technique to both synthetic and real shapes, the latter as a practical technique for streamlining the segmentation of electron tomography (EMT) data. Due to the inherent qualities of EMT data (a large number of complex shapes coupled with low contrast and high noise) extracting structures of interest from tomographic volumes remains a predominantly manual process [1]. For example, the construction of an accurate surface representation for modeling the chick ciliary ganglion (CG) calyx synapse for the study by [2] required the manual tracing of 690 tomographic slices and approximately 6 months (one person) to complete [3]. Segmentation is therefore a significant bottleneck that could potentially be reduced via automated/semi-automated systems informed by machine learning. The shape prior approach described here provides one building block for such systems.

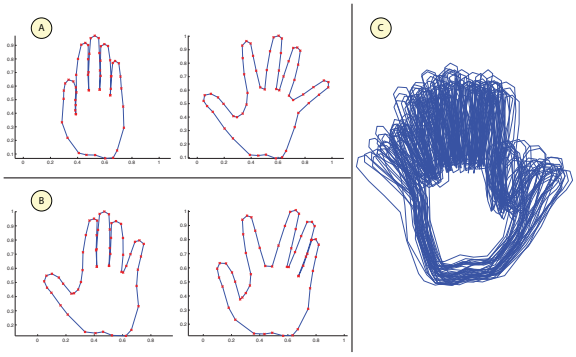

Figure 2. Twelve scanned hand poses, two shown, (A), are algorithmically varied to produce a larger training set (C). Examples of the interpolated shapes are shown in (B).

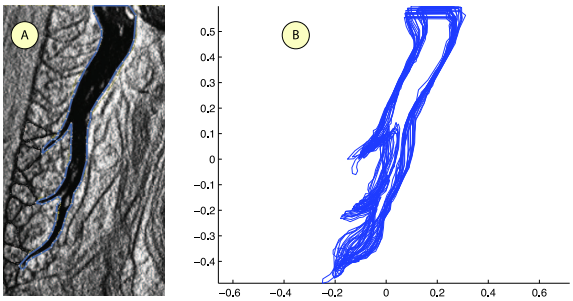

Figure 3. Tracings of 75 tomography slices (A); the resulting myelin shape training set (B).

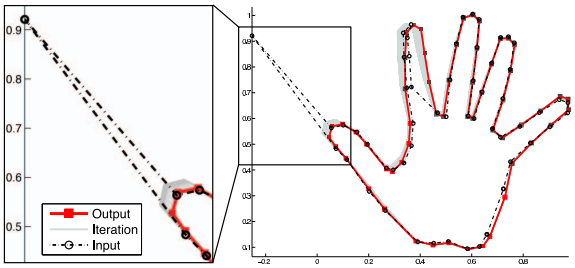

Figure 4. An input shape (dashed black line) is corrupted with one large and three smaller "errors." The shape is projected into a PCA subspace and then adjusted by gradient ascent in probability (grey iterations) producing a plausible shape (red). Please enlarge to see details.

2. RELATED WORK

Shape modeling has an extensive literature [4, 5]. The earliest shape priors were often simple local smoothness preferences such as used in [6]. Global shape eigenmodels became popular in the 1990s [7] and have been widely applied. More recently kernel PCA [8, 9], kernel density modeling [10], and other techniques have been applied in the context of shape prior models.

PCA has several recognized limitations. One is the fact that PCA models only Gaussian statistics, because the data is "seen" through its covariance matrix. The covariance matrix contains the second order moments $E[f(a)f(b)]$ and does not reflect all characteristics of the data. For example, $f(a)f(b) = (-f(a))(-f(b)) = f(b)f(a)$, so the covariance does not capture orientation properties of a signal (Fig. 1).

Despite the limited information present in the covariance, PCA can reconstruct an arbitrary signal to desired accuracy. The remaining information about a particular signal appears in the distribution of the PCA coefficients. With Gaussian data (for which the third and higher-order moments are zero), the PCA coefficients themselves have Gaussian distribution, and plausible synthetic examples of the data can be obtained simply by choosing random coefficients. For non-Gaussian data, the PCA coefficients themselves have some structure that is not characterized by the PC analysis (Fig. 1).

The power of linear techniques such as PCA can be extended by applying a nonlinear mapping to the data before application of the linear technique, thereby exposing new features of the data for analysis. The "kernel trick" further recognizes that it is not necessary to explicitly calculate the eigenvectors of the feature space covariance matrix (a smaller dual eigenproblem is substituted), leading to practical algorithms such as Kernel PCA. Although these techniques bring new capabilities, the kernel corresponding to the nonlinear mapping must be experimentally selected so as to emphasize desired features of the data, and the representation within the feature space is still second order. Alternately, PCA can be extended to less "linear" data by grouping the data into subsets and then applying PCA locally to each subset [11]. Such collections of local models cannot represent coupled global nonlinear characteristics however.

It can be recognized that characterizing a space of shapes is fundamentally a problem of modeling multidimensional probability density. Kernel density approaches [12] can approximate arbitrary probability distributions. On the other hand, it is generally considered that kernel densities do not work well in high dimensions, due to a version of the "curse of dimensionality." The data required to sufficiently fill the space is exponentially proportional to the dimensionality of the data, suggesting that kernel methods have limited applicability in more than several dimensions. The problem is easily explained by considering the proportion of data covered by an n-dimensional kernel of fixed diameter as the dimension is varied. For illustration purposes, suppose that the data is uniformly distributed in an n-dimensional unit ball. In one dimension, a kernel of diameter 0.5 then contains about 50% of the data. In two dimensions, a disc of diameter 1/2 contains 1/4 of the total area, and similarly the volume of an n-dimensional subdisc is dominated by a factor proportional to 0.5^n. Thus, either very large kernels are needed (leading to reduced resolution), or exponentially more data is required to fill the volume.

An evident strategy is to apply kernel density modeling to the PCA coefficients rather than to the original data. This PCA+KD combination offsets the respective limitations of each technique considered separately: the kernel density can model non-Gaussian coefficient structures that are not captured by PCA (Fig. 1), and PCA dimensionality reduction can model useful classes of shapes with 3-5 eigenvectors, thereby partially addressing the curse-of-dimensionality issue in KD modeling. In the next section we demonstrate that practical algorithmic and editing operations can be formulated in terms of PCA+KD shape priors.

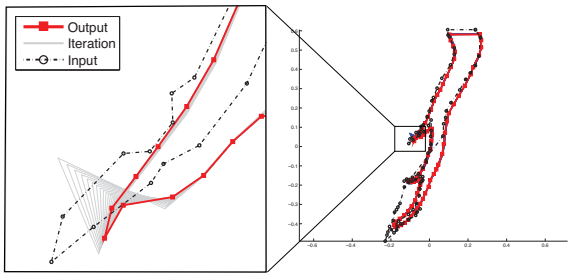

Figure 5. Ten contiguous points on the input contour (dashed black line) are displaced, simulating a segmentation error. The PCA projection produces an implausible self-intersecting shape that is corrected by PCA+KD (grey iterations resulting in red shape).

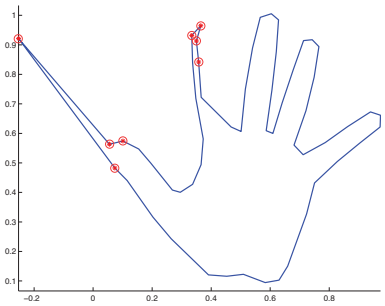

Figure 6. The eight most "implausible" points on this corrupted shape as automatically identified using Eq. 1

.

3. METHODS AND RESULTS

3.1. Data

Our hand shapes training data set was seeded by manually tracing contours of scanned images of a hand in twelve different poses. We then synthetically generated a much larger space of shapes from the initial hand models using linear shape interpolation between pairs of hand shapes, combined with a small amount of correlated random warping to generate realistic variability (Fig. 2). Our myelin shape training data set was obtained by manually tracing seventy-five contours from a series of electron tomography images provided by the National Center for Microscopy and Imaging Research (Fig. 3).

3.2. Identifying and Correcting Noise

In the following we denote the shape as represented in the principal component basis as

$$s = Uc_0 + m$$

where columns of U contain the eigenvectors of the covariance matrix, $c_0 = U^T s_0$ are the PCA coefficients of the original shape s_0, s is the resulting approximated shape, and m is the mean shape. The coefficient distribution is modeled with a 4-dimensional kernel density with width tuned to be roughly 1/4 of the coefficient range.

Figs. 4,5 show the PCA+KD reconstruction of corrupted shapes. In each case, a number of points on a shape (not in the training set) have been displaced to simulate noise or other mistakes in an automatic contour tracing algorithm (dashed line). The corrupted shape is first projected into the PCA subspace, giving a coefficient vector c and a corresponding reconstruction. The reconstructed shape is not ideal (for example, the thumb in Fig. 4 is too long). Gradient ascent in the KD probability is applied to the coefficient vector, resulting in a series of shapes (grey outlines) that converge to an improved shape (red). Note that this global shape evolution is not a realistic application (the shape evolution should be guided by a data fidelity term as well), but it serves to illustrate the capability of the shape prior.

More typical application scenarios require operation on individual points. This requirement arises while tracing or evolving contours, for example; individual points may be iteratively moved so as to minimize an energy that includes data fidelity and prior likelihood terms, and unlikely points may be judged to be outliers and excluded.

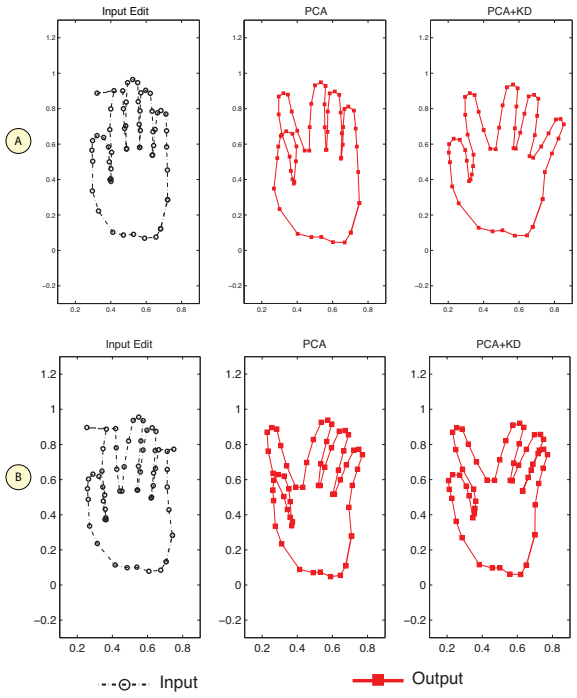

Figure 7. (A, Left) the user adjusts a single point, (middle) the PCA projection has self intersection, (right) PCA+KD removes the self intersection, though it also results in the fingers spreading. The user then runs the process a second time, adding a point constraint to reduce the movement on the little finger (B, left). (Note that slight differences in the index finger area are due to both the differing user input and a different random warping used to expand the training set).

While the marginal probability of individual points can be found by integration over the KD, this is relatively uninteresting because the relation of a point to the configuration of other points is ignored. A better approach to assessing the plausibility of individual points is evident in Fig. 4: unlikely points are those that move the most. This approach assumes that the entire contour is evolving and requires global computation.

A more economical approximate measure of the plausibility of individual points is suggested by the following statement: points that are plausibly situated with respect to the prior will have low probability gradient. Conversely, implausibly situated points will often have a significant probability gradient (although this is not guaranteed). The gradient with respect to change in particular point coordinates can be calculated as

$$\frac{dP(c)}{ds_k} = \frac{dP(c)}{dc}\frac{dc}{ds_k} = \nabla_c P \frac{dc}{ds_k} \tag{1}$$

where s_k is one (x- or y-) coordinate on the shape and $\nabla_c P$ is the probability gradient (with respect to PCA coefficients c) evaluated at c. Since $c = U^T s$, $\frac{dc}{ds_k}$ is $U^T_{:k} = U_{k:}$, the kth row of the eigenbasis. The magnitude of the gradients at a point is then a fast approximate measure of point plausibility (Fig. 6).

3.3. Computer assisted editing

Tomographic segmentation remains a predominantly manual process despite the extant variety of research on automated segmentation techniques [1]. Therefore, a user-guided but computer assisted technique is a more realistic short-term goal than fully automated segmentation. We implement an approach in which the user adjusts the location of one or more points on an initial shape (such as the segmentation result from the previous tomographic slice). The remaining points are automatically evolved toward a plausible shape while interpolating the user-specified "oracle" points.

Coefficients c that interpolate the user-specified oracle points will be used as a starting point for the PCA+KD procedure. There are many possible vectors c; we choose the one that is closest in the L_2 sense to the original c_0. The intuition here is that the new shape will probably have some general similarity to the initial shape, particularly in the case of tomographic slices. This goal can be formulated:

$$\min \frac{1}{2}(c-c_0)^T(c-c_0)+\Lambda^T(Ec-p)$$

where p is a vector containing the oracle points, E contains the rows of U corresponding to the oracle points, and Λ is Lagrange multiplier vector. Setting the derivative with respect to c, Λ to zero gives the system

$$\begin{bmatrix} I & E^T \\ E & 0 \end{bmatrix} \begin{bmatrix} c \\ \Lambda \end{bmatrix} = \begin{bmatrix} c_0 \\ p \end{bmatrix}$$

that provides the desired c.

Figure 7 (A, left) shows a shape with a single oracle point specified by the user (selecting only a single point places the highest demand on the shape prior), and Fig. 7 (A, middle) shows the shape generated using the new coefficients c. The PCA-generated shape at (A, middle) is greatly improved with respect to (A, left), but shows undesirable self-penetration.

The generated shape is improved by appealing to the kernel density. Starting from position c, we walk uphill in probability while continuing to interpolate the oracle points. This can be done by projecting the probability gradient into the nullspace of the constraint matrix E. From the singular value decomposition, $E = USV^T$, form the matrix V_0 consisting of the last $2n$ columns of V for n oracle points. The modified gradient is then

$$\underset{\text{subspace}}{\nabla} P(c) = V_0 {V_0}^T \nabla P(c)$$

Figure 7 (A, right) shows the final computer-assisted shape after crawling uphill to a probability plateau. The self intersection is eliminated and the shape proportions are improved. This example using only one user-specified point shows that the shape prior is effective even in severely underconstrained situations. In panel (B) the user decides to place a second point, constraining the movement of the little finger. This assisted editing process can continue with any desired number of input points, rapidly approximating any desired plausible shape configuration.

4. CONCLUSION

We demonstrated a simple and scalable shape prior modeling method using PCA combined with kernel density. By using PCA to perform dimensionality reduction on high dimensional data, and then modeling selected PCA coefficients using a kernel density, we overcome the second order limitations of PCA while deferring the curse of dimensionality that affects kernel density modeling in high dimensions.

Unfortunately the range of dimensionality that can be considered with this approach cannot be characterized in advance, as it depends on the data. If the PCA coefficients have little local structure, they can be modeled with broad kernels that are required in high dimensions. On the other hand, some complex and highly structured shapes may require high dimensional modeling and a prohibitive amount of training data. In these cases a local rather than global shape model could be adopted. Nevertheless, we show that PCA+KD is effective and improves on PCA on several examples of realistic complexity.

Our results were generated by applying the PCA+KD approach to point distribution shape models of synthetic and biological shapes. Potential applications include shape priors enhanced image segmentation and other applications that could benefit from accurate priors.

5. ACKNOWLEDGEMENTS

This work was in part supported by a National Institutes of Health, PHS P41RR04050D, ME Ellisman, Principal Investigator.

6. REFERENCES

[1] V. Lucic, F. Forster, and W. Baumeister, "Structural studies by electron tomography: From cells to molecules," *Ann. Rev. of Biochemistry*, vol. 74, pp. 833–865, 2005.

[2] J. Coggan, T. M. Bartol, E. Esquenazi, J. R. Stiles, S. Lamont, M. E. Martone, D. K. Berg, M. H. Ellisman, and T. J. Sejnowski, "Evidence for ectopic neurotransmission at a neuronal synapse," *Science*, vol. 309, pp. 446–51, 2005.

[3] T. Bartol and M. E. Martone, "Personal communication," 2006.

[4] C. G. Small, *The Statistical Theory of Shape*, Springer, 1996.

[5] L. Da Fontoura Costa and R. Marcondes Cesar Jr., *Shape Analysis and Classification*, CRC Press, 2000.

[6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proceedings, First International Conference on Computer Vision*, 1987, pp. 259–268.

[7] T.F. Cootes, C. J. Taylor, D.H. Cooper, and J. Graham, "Training models of shape from sets of examples," in *Proc. British Machine Vision Conference*. 1992, pp. 9–18, Springer-Verlag.

[8] D. Cremers, T. Kohlberger, and C. Schnorr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition*, vol. 36, no. 9, pp. –1943, 2003.

[9] S. Dambreville, Y. Rathi, and A. Tannenbaum, "Shape-based approach to robust image segmentation using Kernel PCA," in *CVPR (1)*. 2006, pp. 977–984, IEEE Computer Society.

[10] D. Cremers, S. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: Teaching level sets to walk.," in *DAGM-Symposium*, 2004, pp. 36–44.

[11] Christoph Bregler and Stephen M. Omohundro, "Surface learning with applications to lipreading," in *Advances in Neural Information Processing Systems*. 1994, pp. 43–50, Morgan Kaufmann.

[12] E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.