# Modeling the structure of multivariate manifolds: Shape Maps

Georg Langs, Nikos Paragios
Laboratoire MAS, Ecole Centrale de Paris,
Equipe GALEN, INRIA Saclay, Île-de-France
Grande Voie des Vignes, 92 295 Chatenay-Malabry, France *
georg.langs@ecp.fr, nikos.paragios@ecp.fr

## Abstract

*We propose a shape population metric that reflects the interdependencies between points observed in a set of examples. It provides a notion of topology for shape and appearance models that represents the behavior of individual observations in a metric space, in which distances between points correspond to their joint modeling properties. A Markov chain is learnt using the description lengths of models that describe sub sets of the entire data. The according diffusion map or* shape map *provides for the metric that reflects the behavior of the training population. With this metric functional clustering, deformation- or motion segmentation, sparse sampling and the treatment of outliers can be dealt with in a unified and transparent manner. We report experimental results on synthetic and real world data and compare the framework with existing specialized approaches.*

## 1. Introduction

Models of shape and appearance are powerful tools in various domains. An example are active appearance models [7] that have been used in medical imaging applications [2], or face tracking [11]. Sparse shape models [12] have been used for liver segmentation, and spherical wavelet shape models were introduced in [16] for brain segmentation. This family of methods necessitates an involved and usually supervised learning phase that establishes correspondences over a large set of training data. Due to the complexity of relevant data, recently the automatic learning of models has attracted interest. Examples are the investigation of systematic behavior as opposed to purely elastic deformation in the context of group-wise registration and brain atlases [3, 10], or autonomous model learning [13].

The main limitation of these models relates to the choice of topology and parameterization. During learning and search, they either use an a priori chosen topology to parameterize correspondences and deformation [9], or deform the entire volume [8] in a continuous non-rigid manner.

Figure 1. Data for which a prior choice of a parameterization manifold can introduce a bias: the highly correlated symmetric structure of face data, a stent-graft in the thoracic aorta exhibiting incoherent deformation behavior.

This strategy has difficulties to model certain aspects of natural shape variation, like the presence of discontinuities, articulated structures, high variability of local deformation behavior, or functional properties, that are not reflected properly in a diffeomorphic deformation field. A prior subdivision is not feasible for data with varying elasticity properties, or compound non-rigid structures (soft tissues and implants in anatomical structures). It would neglect the subtle mutual interactions in the data, while a model that does not account at all for the compound nature of the data is likely to suffer from poor generalization behavior [14].

The determination of the intrinsic topology of observations is a necessary prerequisite for the building of models, that are free of a bias, that would empede their analytic value. An example where this is of particular importance are population studies [10]. A related task is the tracking of multiple objects and motion segmentation, which typically relies on assumptions about the behavior of individual components, like rigidity, and aims to derive sub-divisions of trajectory bundles from sequences [18, 21].

*This work has been partially supported by the Region Île-de-France. The authors are grateful to Hicham Kobeiter, Pascal Desgranges, and Alain Rahmouni from Henri Mondor Hospital for providing stent-graft data.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

In summary, shape modeling involves two aspects, 1. the parameterization and 2. the modeling of shape variations. While significant research effort has been carried out towards modeling variations, this is not the case for the parameterization. An a priori choice of parameterization introduces a strong bias in the modeling process. The most natural approach is to determine the parameterization from the data itself to maximize the compactness of the model while simplifying the statistical modeling of shape variations.

In this paper we propose a metric that captures the mutual coherence of landmark behavior from a population of training examples. A *shape map* can be learnt in a simple manner from the data by a construction related to diffusion maps [6, 15]. It is a space in which the joint modeling properties of landmarks are reflected in their Euclidean distance. That is, instead of expressing the relation between individual points of a single observation, it captures the behavior, the interactions, and the inter-dependencies of landmarks in an entire population of observations. Instead of using spatial distances or similarities, it uses the compactness of small sub-models to probe the data for systematic dependencies of groups of landmarks.

This leads to several novel properties that let the shape map serve as exploration instrument, or analysis method that can enhance model building: it establishes functional relations within a deformable structure from multiple observations. By this it captures the individual modeling behavior of sub-model systems independent from their spatial distribution. Shape maps bridge the gap between statistical shape models and diffusion maps. The resulting shape population metric solves several questions in a unified manner: how to discover topology (in terms of dependency) in the data from observations instead of imposing a priori manifolds. How to detect sub-structures, and their mutual relations. How to perform segmentation, determine reconstruction kernels, quantify redundancy, and detect outliers, based on statistical behavior. This is crucial for the statistical modeling, analysis, and reconstruction of complex and compound structures. It enables algorithms to improve model building by focusing on coherent sub-models, and enables the topology to emerge from the observed data.

The reminder of this paper is organized as follows: In Sec. 2 we explain the model compactness measure. Based on this the metric of joint modeling behavior is introduced in Sec. 3. In Sec. 4 several applications and an experimental validation are discussed, while the last Sec. 5 concludes the paper.

## 2. Description Length and Shape Models

We will first explain the information theoretic preliminaries for the learning of a shape map. To probe the landmark set for dependencies we calculate the description lengths of multivariate Gaussian shape variation models, that capture the deformation of landmark sub sets, after they have been aligned by Procrustes analysis.

### 2.1. Modeling Shape Variation

To analyze the model structure, we observe shape models that encompass subsets of the entire data, and model only local non-rigid shape variation. Global translation, rotation or scale change are neglected. Given a set of $m$ landmarks in $n$ $d$-dimensional data examples: $\mathbf{x}_i = \langle\langle x^i_{1,1} \dots x^i_{1,d}\rangle \dots \langle x^i_{m,1} \dots x^i_{m,d}\rangle\rangle$ where $i = 1, \dots, n$, we first align these examples and then calculate the description length of a multivariate Gaussian modeling the remaining non-rigid shape variation. The shape alignment eliminates variability in a set of examples, that is due to translation, rotation and scale. Given $n$ sets of landmarks, $\mathbf{x}_1, \dots, \mathbf{x}_n$, first, each example $\mathbf{x}_i$ is translated so that the centroid is 0, and scaled, so that $\|\mathbf{x}_i\| = 1$. Finally the rotational differences of the sets $\mathbf{x}_2, \dots, \mathbf{x}_n$ are aligned to $\mathbf{x}_1$ (where the choice of $\mathbf{x}_1$ is arbitrary, and does not influence the shape model) using a singular value decomposition of the matrix $\mathbf{x}_i^\top \mathbf{x}_1$ [4], i.e.: $\mathbf{U}\mathbf{D}\mathbf{V}^\top = \mathbf{x}_i^\top \mathbf{x}_1$. Then $\mathbf{V}\mathbf{U}^\top$ is the rotation matrix, that by $\mathbf{x}'_i = \mathbf{V}\mathbf{U}^\top \mathbf{x}_i$ minimizes the Procrustes distance of $\mathbf{x}'_i$ to $\mathbf{x}_1$. By modeling the aligned shapes $\mathbf{x}'_1, \dots, \mathbf{x}'_n$ (where $\mathbf{x}'_1 = \mathbf{x}_1$) only non-rigid shape variation remains, and global translation, scale and rotation do not affect the model complexity. The aligned shapes are modeled by a multivariate Gaussian with model mean $\bar{\mathbf{x}}$ and covariance matrix $\Sigma$. PCA is applied on the set $\{\mathbf{x}'_i, i = 1, \dots, n\}$ creating a new coordinate system that represents each of the vectors by

$$\mathbf{x}'_i = \bar{\mathbf{x}}' + \sum_{j=1}^{n_p} a_j^i \mathbf{e}_j. \tag{1}$$

The modes $\mathbf{e}_j$ are the eigenvectors of the covariance matrix sorted according to decreasing eigenvalue $\lambda_j$. $\bar{\mathbf{x}}'$ is the mean shape and $n_p$ can be chosen to represent a certain amount of variation in the data. The eigenvalues $\lambda_j$ correspond to the variance of the data in the direction $\mathbf{e}_j$, and $a_j^i$ are the coefficients that correspond to one example $\mathbf{x}_i$.

### 2.2. Compactness of multi variate Gaussian models

To quantify the compactness of the shape models capturing the variation of a set of landmarks, we calculate their description length. The description length comprises the cost $L$ of communicating a model $\mathcal{M}$ itself and the data $D$ (i.e. the landmark positions) encoded with the model: $L(D, \mathcal{M}) = L(\mathcal{M}) + L(D|\mathcal{M})$. It can be calculated from the individual univariate Gaussians.

For each dimension $j$ of the eigenspace used to encode the data we can apply Shannons theorem [20] to the according one-dimensional distribution. We quantize the coefficients $a_j^i$ that are strictly bounded by $R_j$, by the step size $\Delta_{Im}$ which is related to the pixel/voxel-size. For

each training sample $\mathbf{x}_i$ the new discrete coordinates $\hat{a}_j^i = k\Delta_{Im}, k \in \mathbb{Z}$ with $-R_j/2 \leq \hat{a}_j \leq R_j/2$ are modeled by a Gaussian distribution with coefficient mean value $\mu_j = 0$ and standard deviation $\sigma_j = \sqrt{\lambda_j}$.

For each dimension $j$ of the eigenspace used to encode the data the transmission costs of the model $L(\mathcal{M}_{\mathbf{e_j}})$ are the quantized eigenvector, $\hat{\sigma_j}$ and the quantization parameter $\delta_j$ for the direction $\mathbf{e}_j$. $L(D|\mathcal{M}_{\mathbf{e_j}})$ is the cost of transmitting the data i.e. the quantized coefficients $\hat{a}_j^i$ of the training set with respect to the direction $\mathbf{e}_j$.

The description length for the data encoded with an $n_p$ dimensional eigenspace is the sum of the transmission costs for the data encoded using the eigenvectors $(\mathbf{e}_j)_{j=1,\ldots,n_p}$ together with the cost of the residual error

$$\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{j=1}^{n_p} \left( L(\mathcal{M}_{\mathbf{e_j}}) + L(D|\mathcal{M}_{\mathbf{e_j}}) \right) + \mathcal{R}, \quad (2)$$

where

$$L(\mathcal{M}_{\mathbf{e_j}}) + L(D|\mathcal{M}_{\mathbf{e_j}}) = \log_2\left(\frac{\sigma_{max} - \sigma_{min}}{\delta_j}\right) + 1 + \quad (3)$$

$$+|\log_2 \delta_j| - n \log_2 \Delta_{Im} + \frac{n}{2}\log_2(2\pi\sigma_j^2) + \frac{n}{2} + \frac{n\delta_j^2}{12\sigma_j^2}, \quad (4)$$

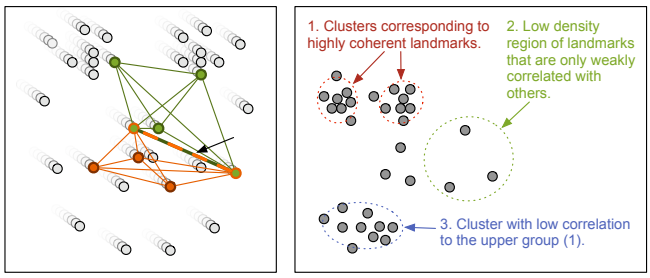
with $\sigma_{max} = R/2$ and $\sigma_{min} = 2\Delta_{Im}$. $\mathcal{R}$ is the residual error that remains after fitting the training set with the model. An extensive derivation of the description length calculation for Gaussian models is given in [9]. Note that the description length can be applied to any model, and a multivariate Gaussian is chosen as an illustrative example. The description length reflects the complexity of the representation, and therefore we can use it in the following step, to describe the coherence of behavior for sets of landmarks.

## 3. Markov Chains and Shape Maps

We will first explain the steps necessary for the construction of the shape map, and will then outline its properties.

Given a set of $n$ example shapes, each consisting of positions for $m$ landmarks as described in Sec. 2.1, we want to derive a metric on the set of landmarks, that reflects their joint modeling behavior, or the coherence of deformation of groups of landmarks. To do this we first learn a Markov chain, that captures these relations between pairs of landmarks, using the description length to evaluate the landmark coherence. From this we derive the *shape map*, in which each landmark is mapped to a point, and in which the Euclidean distance between points relates to the complexity of their optimal joint model. The temporal ordering of examples or connectivity is neglected entirely in this process. The most related work is [6] in which *diffusion maps* have been explained in detail.

The Markov chain consists of $m$ nodes $X$, and pairwise relations $d(i,j)$. Each node corresponds to one landmark, and therefore *carries* information about $n$ instances,




**Generating the Markov chain:** Each node carries landmark position information for multiple examples. To calculate the weight of the edge d(i,j) (indicated by the arrow), we probe the deformable shape with multiple sub-models.

**Resulting shape map:** Each landmark is mapped to a position mirroring its inter-dependencies (joint modeling behavior) with other landmarks. The closer two landmarks, the more coherent their behavior.

Figure 2. Learning the Markov chain, and the resulting shape map.

A model family $M = \langle \mathcal{M}, \Phi \rangle$ parameterized by $\Phi$ allows us to formulate a notion of *systematic* shape variation with regard to this family. This mutual systematic dependency serves as a measure of affinity between landmark behavior.

**A kernel describing systematic behavior** We define $d_k(i,j)$ for two landmarks and *kernel size* $k$ based on the minimum description length of models encompassing the two landmarks $i$ and $j$ and $k-2$ other landmarks:

$$d_k : \{1, 2, \ldots, n\}^2 \to \mathbb{R} \quad (5)$$

$$d_k(i,j) = \min_{\mathcal{M}}(\mathcal{L}(\mathcal{M}) | i,j \subseteq \mathcal{M} \text{ and } \#\mathcal{M} = k) \quad (6)$$

where the model $\mathcal{M}$ with cardinality $\#\mathcal{M} := \#\{h_1, \ldots, h_k\} = k$ represents the sub vectors comprising a sub set of landmarks of the observations $(\mathbf{x}_i^{\{h_1,\ldots,h_k\}})_{i=1,\ldots,n}$, where $\mathbf{x}_i^{\{h_1,\ldots,h_k\}} = \langle\langle x_{h_1,1}^i \ldots x_{h_1,d}^i\rangle \ldots \langle x_{h_k,1}^i \ldots x_{h_k,d}^i\rangle\rangle$. Accordingly $\mathcal{L}(\mathcal{M})$ denotes the description length $\mathcal{L}((\mathbf{x}_i^{\{h_1,\ldots,h_k\}})_{i=1,\ldots,n})$ as defined in Sec. 2.2, and $i,j \subseteq \mathcal{M} :\Leftrightarrow i,j \in \{h_1, \ldots, h_k\}$ for $\mathcal{M}$. That is, $d(i,j)$ is the minimum of the description lengths of all models representing $i$ and $j$, and arbitrary $k-2$ other entries of the observations. Note that therefore $d_k(i,j)$ does not only depend on the behavior of the two landmarks $i$ and $j$, but on the behavior of a larger sub set that can be described by the least complex joint model. With increasing $k$ only larger coherent sub-sets will benefit in terms of the distance in the shape map. This can have a positive impact on the robustness. $d_k$ is non-negative $d_k(i,j) \geq 0$, and symmetric i.e. $d_k(i,j) = d_k(j,i)$. That is the nodes $X$ and edges weighted according to $d_k$ between the nodes build a symmetric graph $(X, d_k)$. In practice we estimate $d_k$ by randomly choosing sub-sets of the data, calculating the according value $d_k(i,j)$ for all pairs in the sub-set and keeping the minimum of all samples for $(i,j)$ (Fig. 2).

**Creating a Markov chain that reflects the model structure** Using this relation, one can form a Markov chain that

encompasses the notion of compactness in the entire set of landmarks. The normalized graph Laplacian construction [5] allows us to construct a reversible Markov chain from the kernels

$$k(i,j) = e^{-\frac{d_k(i,j)}{\epsilon}} \tag{7}$$

by defining

$$d(i) = \sum_j k(i,j) \quad \text{and} \quad p(i,j) = \frac{k(i,j)}{d(i)}. \tag{8}$$

This new kernel is no longer symmetric, but satisfies

$$\sum_j p(i,j) = 1. \tag{9}$$

Therefore it can be interpreted as the probability of the transition from node i to node j in one time step, or a *transition kernel* of a Markov chain. It gives a diffusion operator

$$Pf(x) = \sum a(x,y) f(y) d\mu(y), \tag{10}$$

and its powers $P^t$ that allow to propagate information through the Markov chain according to the transition kernels. $P$ is the Markov matrix with the entries $p(i,j)$. The probability of the transition from any node $i$ to another node $j$ in $t$ timesteps is given by the according kernel $p_t(i,j)$. This allows to analyze the data at multiple *scales* i.e to propagate the relations between pairs of nodes.

**A distance that captures coherence** The operator $P$ defines the geometry on the set of landmarks we are looking for. It can be mapped to an Euclidean geometry by an eigenvalue decomposition of $P$. According to $P^t$ we can define a family of *diffusion distances* parameterized by $t$ on the set of landmarks

$$D_t(i,j) = \sum_{l=1,\ldots,m} \frac{(p_t(i,l) - p_t(j,l))^2}{\pi(l)} \tag{11}$$

where $\pi(i) = d(i)/\sum_j d(j)$ is the probability of $i$ in the unique stationary distribution (the uniqueness is fulfilled if the graph is connected). $D_t$ is an $L^2$ distance between the posterior distributions of reaching $i$ or $j$ from all points $l$ in the graph. It captures the connectivity in the Markov chain, summing over all possible paths from $i$ to $j$. The distance $D_t$ is low if there is a large number of paths of length $t$ with high transition probabilities between the nodes $i$ and $j$. In terms of the joint model this signifies high compactness for a representation that encompasses landmarks $i$ and $j$.

An eigenvalue decomposition of the operator $P$ results in a sequence of eigen values $\lambda_1, \lambda_2 \ldots$ and corresponding eigen functions $\Psi_1, \Psi_2, \ldots$ that fullfill $P\Psi_i = \lambda_i \Psi_i$. In [6] the authors explain how a *diffusion map*, the space spanned by the eigenfunctions of a Markov chain relates to the geometry determined by a diffusion distance $D_t$.

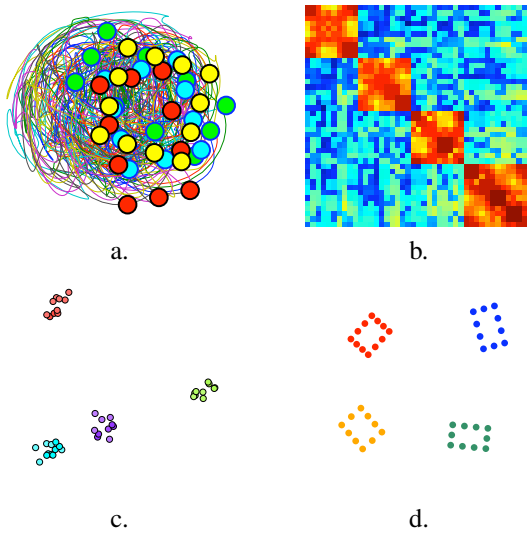



Figure 3. Synthetic example: a. trajectories of 4 boxes, rotating against each other, while changing aspect ratio individually; b. estimated Markov matrix; c. clusters in the space map; d. landmarks corresponding to the 4 clusters: separation of the boxes

**Shape maps: a geometry of modeling behavior** We use this to create a *shape map* $\mathbf{\Psi}_t : X \rightarrow \mathbb{R}^w$, that embeds each node (landmark) $i = 1, \ldots, m$ in the Markov chain into a $w$ dimensional Euclidean space

$$i \mapsto \mathbf{\Psi}_t(i) \triangleq \begin{pmatrix} \lambda_1^t \Psi_1(i) \\ \lambda_2^t \Psi_2(i) \\ \vdots \\ \lambda_w^t \Psi_w(i) \end{pmatrix} \tag{12}$$

in which the diffusion distance in Eq. 11 corresponds to the euclidean distance

$$\|\mathbf{\Psi}_t(i) - \mathbf{\Psi}_t(j)\| = D_t(i,j). \tag{13}$$

Thereby the *functional* relations between landmarks are translated into spatial distances in the shape map (Fig. 2).

We can perform standard density estimation, clustering, or the definition of neighborhoods in this space. Such a process captures relations between the behavior of landmarks in the population of examples, since the underlying Markov chain was built by description length kernels.

The diffusion distance $D_t$ is able to derive large scale relations based on pairwise connections in the Markov chain, by using higher values for $t$. This property can be exploited when estimating the Markov chain on large structures. When only spatial neighbors are used for the pairwise relations, the relations between landmarks that are not captured in this calculation can be estimated through the propagation when running the Markov chain, i.e. increasing $t$. The time in the Markov chain corresponds to a simple scaling of the shape map dimensions. It allows us to handle

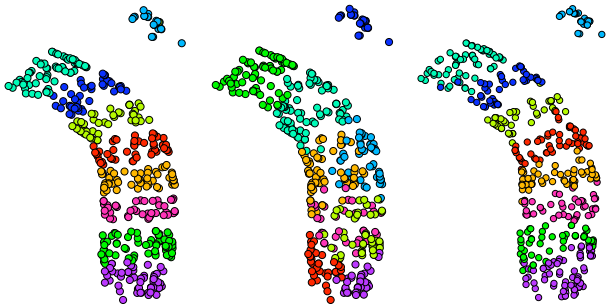



Figure 4. Stent-graft segmentation: a. ground truth, b. spatial segmentation, c. deformation segmentation with shape maps.

a certain amount of non-linear behavior of groups of landmarks even if we just use a linear model for the pair-wise relations. In our experiments we did not consider spatial neighborhoods during the estimation process, in order to avoid its influence in judging the value of the purely functional relations between landmark behavior.

## 4. Using Shape Maps - Experiments

The shape map is a metric space in which we can perform clustering, and density measurements. Furthermore, we can use the neighorhood relations to guide operations in the real population, like the reconstruction of missing values, for new examples. We will illustrate these aspects and applications of the shape map metric in experiments on synthetic and real world data sets from various domains: 1. a synthetic set of landmarks on 4 rectangles, each of them changing aspect ratio, while they rotate against each other. 2. A sequence of landmarks on a talking face. 3. Frames of a gated CT sequence of a stent-graft in the thoracic aorta. 4. Two motion segmentation data sets with rigid motions and outliers [17]. Spatial relations in the data were neglected during the shape map estimation.

### 4.1. Deformation segmentation

We segment the set of landmarks, accounting for its shape variation behavior. The segmentation separates sub structures that have high intra-set coherence, while being comparably independent from other parts of the data. This can be realized by spatial clustering in the shape map. We performed k-means clustering in the shape map for the experiments. The cluster centers in the shape map correspond to points that have equal probability of being generated by the same model, like all other members of the cluster.

In Fig. 3.a a synthetic example is depicted. It consists of 75 examples of four boxes that rotate independently from each other, while changing their individual aspect ratios. The Markov matrix in Fig. 3.b reflects the dependencies

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 31 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 22 |
| 0 | 45 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 29 | 0 | 0 | 0 | 16 | 23 | 0 |
| 0 | 0 | 0 | 39 | 0 | 28 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 0 |
| 0 | 21 | 0 | 9 | 0 | 14 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 54 | 0 | 0 |
| 8 | 0 | 19 | 0 | 0 | 0 | 0 | 15 | 0 |
| 20 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 27 |

Spatial segmentation

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 0 | 43 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |
| 0 | 0 | 68 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 56 | 0 | 8 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 43 | 0 | 0 | 0 |
| 0 | 0 | 6 | 0 | 0 | 0 | 46 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 46 |

Shape map segmentation

Table 1. Confusion matrices for 8 stent segments and a static part of the spine.

within the boxes, and the resulting shape map exhibits four compact clusters shown in Fig. 3.c. They correspond to the landmarks located on the four individual boxes, as depicted in Fig. 3.d.

A *stent-graft* in the thoracic aorta was tracked in a sequence of 8 frames acquired during the cardiac cycle. During the sequence the entire stent-graft deforms non-rigidly, and changes topology due to individual segments touching each other either during the entire sequence or only for a few frames. In Fig. 1 one example is depicted. For 451 landmarks positions in all frames are known. Landmarks were located on 8 stent segments and a close static structure on the spine, therefore we expected 9 separable sub-structures in the data. The ground truth segmentation was annotated manually for validation. Spatial and shape map clustering was performed. The spatial clustering resulted in a success rate of $62\%$, while the shape map clustering resulted in a success rate of $94\%$. In Fig. 4.a the ground truth annotation, and the results from spatial (Fig. 4.b) and shape map clustering (Fig. 4.c) are displayed for 9 clusters. In Tab. 1 quantitative results are reported. The spatial segmentation does not result in usable segments, while the shape map segmentation separates the individual - and less elastic parts - of the stent-graft.

### 4.2. Density estimation and spatial relations

We can estimate the density in the shape map by a Gaussian kernel. It corresponds to the quantity and affinity of landmarks that exhibit coherent behavior in the population, in terms of that a model that captures this set is compact. Therefore the density can by used for a determination of the number of clusters, by enforing positions of cluster centers

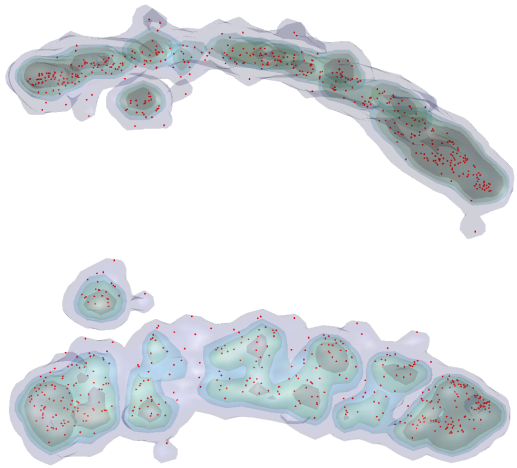


Figure 5. Stent-graft deformation: density in the shape map: high density regions correspond to landmarks of individual stent segments that exhibit coherent deformation behavior. The isolated cluster corresponds to landmarks not located on the stent-graft, but on a static part of the spine.

in high density regions of the shape map. On the other hand the local density in the shape map gives a means of confidence of cluster membership, and allows for the detection of outliers, by rejecting points in low density regions. In Fig. 5 two views of the density in a three-dimensional shape map of the stent-graft data is depicted. It shows the dense clusters formed by the 8 segments and the static part of the spine. Pairs of segments at both ends of the stent-graft form narrow cluster pairs, indicating their less elastic connection. Note that the structure forms a rather flat two-dimensional manifold.

### 4.3. Modeling, reconstruction and tracking

Furthermore the shape map and its density distribution allows for a reduction of the shape representation dimensionality. Utilizing landmarks corresponding to evenly distributed nodes in the shape map reduces the redundancy in the landmark set, since only few examples are necessary to represent dense areas, that correspond to highly redundant behavior in the training set. When building a model or reconstructing missing data in new observations, the metric in the shape map gives a neighborhood relation, that can be used for reconstruction based on local neighborhoods in the shape map i.e. from landmarks, that are most related in terms of behavior. This can be viewed as a notion of topology emerging from the behavior of observed data. It is valuable in work concerned with the autonomous building of models, where the correct determination of the underlying structure that generates the observation is a prerequisite for accurate models with good generalization behavior.

Landmarks on the face of a talking person were tracked

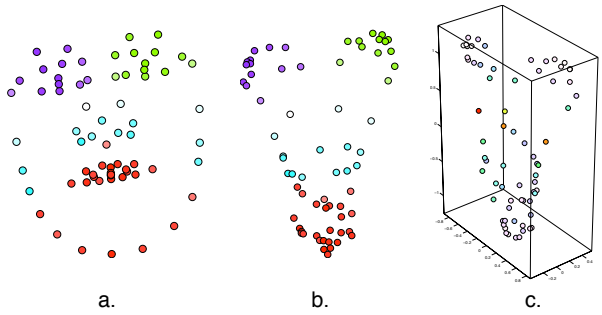



Figure 6. Modelling of a face based on 75 randomly chosen frames from a sequence of a talking person. Colors correspond to submodels. In a. and b. the saturation encodes the density in the shape map. High saturation corresponds to high density. Left: original space, center: shape map. On the right the density is color coded on images of the landmarks in the first 3 shape map dimensions.

(Fig. 1, [1]), and 75 frames were randomly chosen from the sequence to test the algorithm. In Fig. 6.b the resulting shape map is depicted together with the mean shape (Fig. 6.a). Clustering was performed dividing the face into 4 zones. Note that spatial relations are not used during the shape map calculation. The density in the space map is visualized by the saturation of the cluster labels. In Fig. 6.c the first 3 dimensions of the shape map are shown, illustrating the structure that reflects the shape variation. Highly correlated clusters are located at the eyes, and the mouth. For four landmarks in the center of the face the coherence with other parts is particularly low.

To assess the the value of shape maps for the reconstruction of missing data, we performed imputation [19] of missing landmarks in the box data set. 34 examples were used for training. In the remaining examples missing landmarks were reconstructed from the other landmark positions in a leave-one-out manner. The configuration diameter is appr. 80px. Reconstruction with the full covariance matrix fails due to the non-linear displacements, and yields a mean position error of 17.46px. Local reconstruction from a subset of 5 closest spatial neighbors yields a mean error of 14.90px. When one instead uses the 5 closest neighbors in the shape map, the reconstruction error is reduced to 2.46px, since this allows to use the most affine landmarks in terms of modeling behavoir, which can not be derived from spatial information.

### 4.4. Motion segmentation and point matching

To evaluate the applicability of shape maps for motion segmentation, we performed experiments on the *desk* and *office* data sets in [17], depicted in Fig. 7 and Fig. 8. For both data sets the annotation, and trajectories are depicted in the top frame. The shape map segmentation result is depicted in the center frame, with colors indicating cluster identity and saturation encoding the density in the shape

| | O1 | O2 | O3 | o/l |
|---|---|---|---|---|
| 1 | 49 | 0 | 0 | 1 |
| 2 | 0 | 50 | 0 | 0 |
| 3 | 0 | 0 | 48 | 1 |
| o/l | 4 | 2 | 11 | 34 |

| | O1 | O2 | O3 | o/l |
|---|---|---|---|---|
| 1 | 113 | 0 | 0 | 0 |
| 2 | 0 | 85 | 0 | 0 |
| 3 | 0 | 0 | 76 | 2 |
| o/l | 7 | 9 | 1 | 14 |

Table 2. Confusion matrix for motion segmentation of 3 different objects with outliers: left: *desk data*, right: *office data*.

map for each landmark. The lower frame shows the clusters in the shape map. For two images showing a desk from different viewpoints, 149 corresponding points were annotated on 3 different independently moving objects, and 51 outliers with incorrect correspondences were added. Shape maps with $k = 3$ and $\epsilon = 1500$ were built for both the set without outliers, and the set with outliers. Outliers were detected based on the density in the shape map. Landmarks with a density less than one standard deviation below mean density were rejected. On the set without outliers, clustering in the shape map results in correct separation between the 3 objects. For data with outliers the resulting confusion matrix is given in Tab. 2 (left). Note that there is no confusion within the objects, but only between objects and outliers. On a second data set *office* (Fig. 8) 3 objects and outliers were segmented with shape maps with the same parameters, resulting in the confusion matrix in Tab. 2 (right). Again, the 3 objects are segmented correctly, while outliers are included occasionally. For both data sets no task specific model was used (like e.g. rigid motion). The shape map was built based on multivariate Gaussian sub-models like in the previous examples. The Gaussian model, and its *tolerance* towards deviation from a rigid motion are suspected to be the reason for the inclusion of outliers, that deviate only slightly from the rigid motion. However, the local density in the shape map proves to be a good indicator of outliers with regard to shape behavior.

## 5. Conclusion

We propose a metric that reflects the shape modeling behavior of observed landmarks in a population of examples. Shape maps are learnt based on a Markov chain that captures the complexity of sub-models in the data, and determines modeling relations between landmarks. The resulting metric allows to perform tasks like deformation or motion segmentation, the determination of redundancy, the rejection of outliers, and the use of functional affinity in a simple Euclidean frame work. The description length based learning of the shape map can be applied to any modeling framework. The work is aimed at understanding the fine and often subtle *granularity* of shape and model behavior. The neighborhood relations in the shape map reflect the intrinsic geometry of the data behavior and can support the autonomous building of models, where the proper use of underlying structures is essential for accurate and representative models. The potential of the approach is illustrated

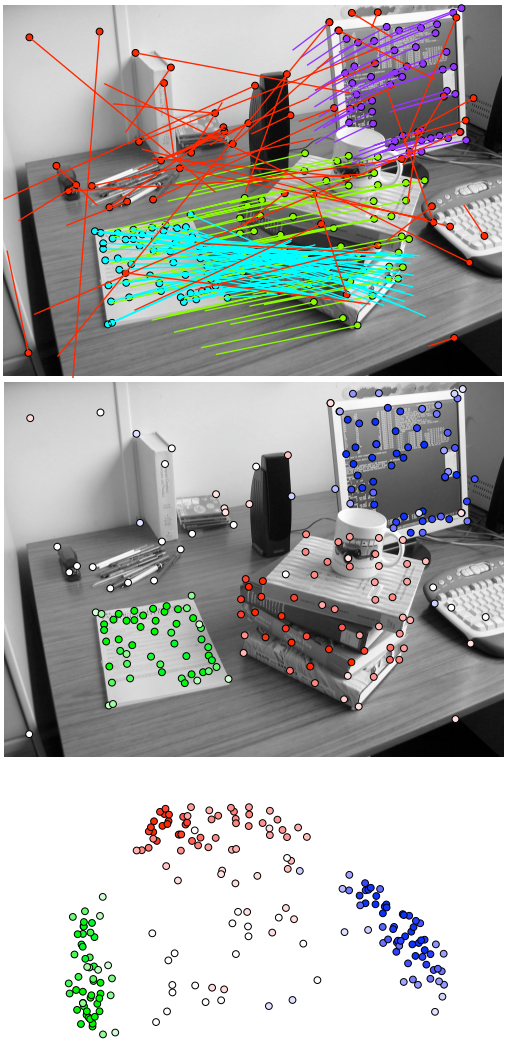


Figure 7. Motion segmentation of rigid objects with outliers from 2 images from different viewpoints *desk data*: annotation and trajectories; automatically labeled motion components and density in the eigenspace, used for outlier detection; clusters in the eigenspace. Data from [17].

on examples from different domains. Future work will focus on the study of non-linear manifolds in the shape map, to further reduce the complexity of the representation, and on the application to different domains, like brain activity, where the description length based approach can provide for a more specific structure representation, than correlation measures.

## References


[1] FGnet - IST-2000-26434, Face and Gesture Recognition Working Group, 2000.

[2] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE TMI*, 24(9):1151–1169, 2005.

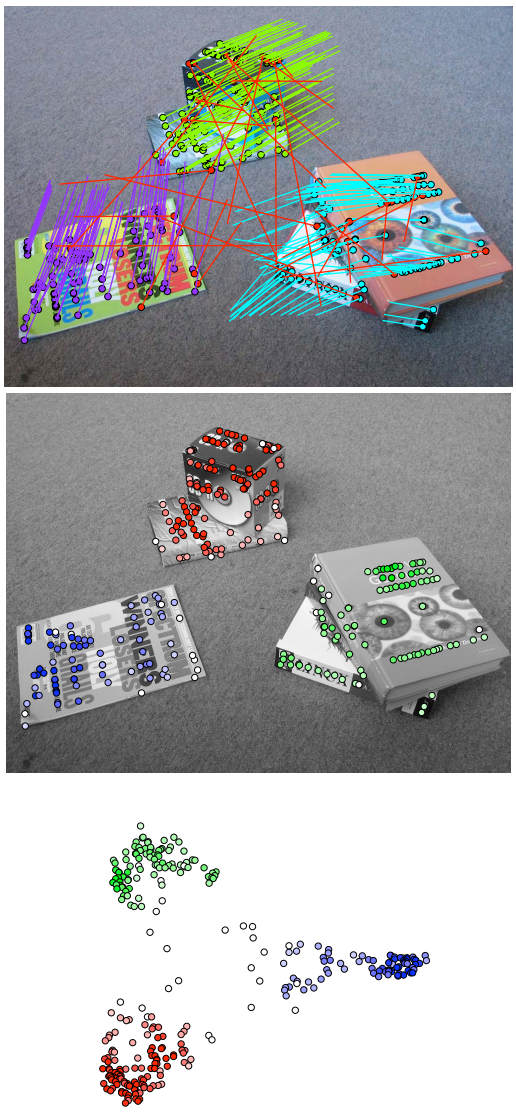


Figure 8. Motion segmentation of rigid objects analogous to Fig. 7 for *office data* [17].


[3] K. Bhatia, J. Hajnal, B. Puri, A. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *Proc. IEEE Intl-Symp. Biomedical Imaging: Nano to Macro,*, volume 1, pages 908– 911, 2004.

[4] F. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

[5] F. R. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[6] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21:5–30, 2006.

[7] T. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE TPAMI*, 23(6):681–685, 2001.

[8] T. Cootes, S. Marsland, C. Twining, K. Smith, and C. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *Proc. of ECCV 04*, pages pp. 316–327, 2004.

[9] R. H. Davies, C. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE TMI*, 21(5):525–537, May 2002.

[10] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi. Population shape regression from random design data. In *Proc. ICCV'07*, 2007.

[11] F. Dornaika and J. Ahlberg. Fast and reliable active appearance model search for 3-d face tracking. *IEEE Trans. on Systems, Man and Cybernetics - Part B Cybernetics*, 34(4):1838–1853, August 2004.

[12] C. Florin, N. Paragios, G. Funka-Lea, and J. Williams. Liver segmentation using sparse 3d prior models with optimal data support. In *Proc. IPMI'07*, volume 20, pages 38–49, 2007.

[13] G. Langs, R. Donner, P. Peloschek, and H. Bischof. Robust autonomous model learning from 2d and 3d data sets. In *Proc. of MICCAI'07*, 2007.

[14] G. Langs, P. Peloschek, R. Donner, and H. Bischof. Multiple appearance models. *Pattern Recognition*, 40(9):2485–2495, 2007.

[15] F. Meyer. Learning and predicting brain dynamics from fmri: a spectral approach. In *Wavelet XII conference, Proceedings of SPIE Volume 6701*, 2007.

[16] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3-d shape representation and segmentation using spherical wavelets. *IEEE Trans Med Imaging*, 26(4):598–618, 2007.

[17] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *IEEE TPAMI*, 28(6), 2006.

[18] K. Schindler, J. U, and H. Wang. Perspective n-viwe multipbody structure-and-motion through model selection. In *Proc. ECCV'06*, 2006.

[19] T. Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate*, 14:853 – 871, 2001.

[20] C. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 1948.

[21] J. Yan and M. Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *Proc. ECCV'06*, volume 4, pages 94–106, 2006.