Statistical Shape Spaces for 3D Data: A Review

Alan Brunton, Augusto Salazar, Timo Bolkart, Stefanie Wuhrer

Chapter 1

Statistical Shape Spaces for 3D Data:
A Review

Alan Brunton[∗], Augusto Salazar[†], Timo Bolkart[‡] and Stefanie Wuhrer[§]

Fraunhofer Institute for Compute Graphics Research IGD
Competence Center 3D Printing Technology
Fraunhoferstrasse 5,
64283 Darmstadt, Germany
alan.brunton@igd.fraunhofer.de

[∗]Fraunhofer IGD, Corresponding author.
[†]Instituto Tecnológico Metropolitano and Universidad de Antioquia, Medellín, Colombia
[‡]Saarland University, Germany
[§]INRIA Rhône-Alpes, France

2   *A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer*

> Methods and systems for capturing 3D geometry are becoming increasingly commonplace–and with them a plethora of 3D data. Much of this data is unfortunately corrupted by noise, missing data, occlusions or other outliers. However, when we are interested in the shape of a particular class of objects, such as human faces or bodies, we can use machine learning techniques, applied to clean, registered databases of these shapes, to make sense of raw 3D point clouds or other data. This has applications ranging from virtual change rooms to motion and gait analysis to surgical planning depending on the type of shape. In this chapter, we give an overview of these techniques, a brief review of the literature, and comparative evaluation of two such *shape spaces* for human faces.

1. Introduction

In recent years it has become evermore affordable and practical to acquire 3D data–through structured light depth sensors, binocular and multi-view stereo systems, and laser scanning systems. This has created an abundance of 3D data, including databases of real examples of 3D shapes of particular classes.

Such databases allow the use of machine learning techniques to automatically extract representations and statistical properties of the 3D shape of these object classes. Having learned such a *statistical shape model* for an object, we can use it to predict the shape of future instances of that object. Put another way, we can learn a *statistical shape space*, in which we expect future observations of that object to lie.

This in turn makes it more practical to find and extract clean, and even recognize instances of the object in raw input data, which may be corrupted by noise, missing data, and even occlusions.

Of particular interest in many applications are the shape of human faces, human bodies, and human organs and skeletal structures. Statistical shape models for these classes of objects, along with algorithms to fit them to ambiguous input data, are important for human computer interaction (e.g. face and expression recognition, gesture recognition, virtual change rooms), surveillance (e.g. gait analysis, sex recognition, face recognition) and telepresence (e.g. virtual avatars), and segmenting organ shapes for diagnosis and surgical planning.

This chapter is based on an article by the same authors,[1] which contains more details and an extensive comparative evaluation. This chapter is intended as a tutorial-style review and explanation of statistical shape space. The learned statistical models described in this chapter are available online.[2]

2. Statistical Shape Spaces

When analyzing 3D data, finding the right space to do it in can make life much easier. For example, having a compact basis, which spans a large subset of possible instances of a complex shape, such as a human, makes it possible to generate instances of that shape from a small set of shape parameters. This allows the shape

parameters to be optimized so that the generated object matches some ambiguous input data, while subjecting the shape parameters to constraints derived from training data.

In this vein, we formulate statistical shape analysis of a given object class as the task of finding a *statistical shape space* that efficiently and informatively represents the shape of objects of that class. We define a statistical shape space as a *shape space* equipped with a probability distribution, or *prior*, measuring how likely it is that an object of the given class would have a particular parametric representation in the shape space. The shape space itself is defined by the set of coefficients obtained by projecting the shapes onto the set of basis functions. (We use the terms basis functions and basis vectors somewhat interchangeably in the following; strictly speaking a basis function is only relevant for continous surfaces, and in practice basis vectors are used for discrete data.)

Thus, we focus on statistical shape analysis as a *generative* technique. A surface containing n vertices in $\mathbb{R}^3$ is represented by d shape parameters or coefficients, which form a vector $\mathbf{s} \in \mathbb{R}^d$. A generator function

$$\mathbf{F}(\mathbf{s}) : \mathbb{R}^d \to \mathbb{R}^{3n} \tag{1}$$

generates from these shape parameters a surface representation (either mesh or point cloud) of n vertices. These shape parameters, and by extension the surface, can be fit to input data of varying modalities (3D point clouds, 3D voxel images, 2.5D depth, 2D images, sparse measurements, etc.), so long as there is a way to measure the distance between the surface and the data, or the quality of the fitting.

As we see in the following review, by far the most common form of statistical analysis used for shapes is principle component analysis (PCA), which seeks a basis in which variance of the training data is maximized. The resulting basis vectors are the directions of greatest variation within the training data. Projecting the training samples onto this basis results in a diagonal sample covariance matrix. If the underlying distribution of the data is assumed to be multi-dimensional Gaussian, then this corresponds to the maximum likelihood estimate of the parameters of the density function. As a result, the resulting shape space is often equipped with a Gaussian prior. If this assumption does not hold, then a Gaussian prior may be arbitrarily far from the true prior, and choosing the correct prior may be challenging. A mathematical description of how to learn a statistical shape space using PCA is given in Section 3.2.

The main difference between statistical shape spaces is the choice of space where PCA is performed, or how the shapes are transformed before applying PCA. This choice leads to different properties, both statistical and computational, of the resulting model. For many models proposed in the literature, the transformation is a change of basis, and the basis of the resulting shape space is composed of the transformed basis and subspace PCA bases (directions of largest variation in each subspace). For other models, a nonlinear transformation is applied followed by a global PCA.

4 *A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer*

Table 1.   Organization of statistical shape fitting methods reviewed.

Type of data	Influence of basis functions	Methods	Artic. separate	Simul. param.
Faces				
	global	Morphable model (PCA)[3–6]	−	−
	global	Morphable model (PCA)[7]	√	−
	global	Multilinear model[8–11]	√	−
	part-based	Part-based model[12–14]	−	−
	part-based	Part-based model[15]	√	−
	localized detail	Hierarchical pyramids[16]	−	−
	local	Local wavelet model[17]	−	−
	local	Multilinear wavelets[18]	√	−
Bodies				
	global	PCA model[19–23]	−	−
	global	SCAPE model[24–29]	√	−
	global	SCAPE model[30]	√	√
	global	Rotation-invariant encoding[31,32]	√	−
	global	Multilinear model[33]	√	−
	global	Posture-invariant model[34]	−	−
	part-based	Segmented PCA model[35]	−	−
	part-based	Part-based multilinear model[36,37]	√	−
Medical Data				
	global	Active shape model (PCA)[38,39]	−	−
	global	Active shape model (PCA)[40]	−	√
	global	PGA model[41]	−	−
	part-based	Part-based model[42]	−	−
	part-based	Part-based multilinear model[43]	√	−
	local	Local wavelet model[44–49]	−	−

This section reviews work on performing statistical shape analysis of 3D data for image processing applications. The categorization of the statistical models in different application domains is summarized in Table 1, where models and methods are grouped by the type of data they were applied to and by the extent of the basis functions used. The table further contains information on whether the models were designed to analyze articulation variations separately from shape variations (here, articulation can refer to facial expression, body posture, or the pose of bones before and after an operation).

In this study, we focus on shape variations over a sample from a population, and hence in Sections 3.2, 3.3 and we analyze and compare statistical models for faces without expression variations. In Section 5 we perform an extensive experimental comparison of the models. This allows us to better examine the differences between the statistical models themselves with respect to the model-fitting task.

The first step to performing shape analysis is to acquire and register a set of *training shapes* that capture the shape variability that is of interest for a particular application. Subsequently, statistical shape analysis is performed on the registered

training shapes: the shapes are projected onto a basis of choice and a probability distribution is fitted to the resulting coefficients to obtain a prior distribution for the shapes of interest. Without correspondence information, this statistical analysis is not possible. However, as indicated in the last column of Table 1, a few methods simultaneously compute a parametrization of a population of shapes while building a statistical model.

Computing correspondences between a population of shapes is a challenging problem, and a detailed discussion about possible approaches is beyond the scope of this work. We refer the reader to recent surveys[50,51] for more information. However, we emphasize that the quality of the registration greatly affects the quality of the resulting statistical models, and by using a high-quality registration in this study, computed as discussed in Section 5.1, we are able to better analyze the properties of models themselves rather than the effects of gross mis-registration.

In computer vision, statistical 3D shape models are commonly used to infer the three-dimensional shape of an object from images, mostly for the purpose of image manipulation. While recently, different classes of shapes have been considered,[52,53] shape models of human faces and human bodies are of special interest due to their immense applicability in human–machine interaction. In medical image analysis, statistical shape models are commonly used to segment medical images and to find correspondences and abnormalities of anatomical shapes. In the following, we review statistical shape spaces used to analyze human faces (Section 2.1), human bodies (Section 2.2) and medical data (Section 2.3).

2.1. *Human Face Shapes*

Blanz and Vetter[3] pioneered the use of statistical modeling for 3D human face shapes with the *morphable model*, which performs a PCA of both shape and texture data from registered 3D face scans and allows the fitting of a 3D face shape to a single image by searching in the learned shape+texture space. While texture is an important cue for human faces, we consider only shape models here. The success of this model spurred many extensions and improvements including: part-based models, both manually segmented[3,12,13,15] and automatically;[14] varying expression;[7] use as a strong prior for multi-view stereo;[4] and distinctiveness preserving priors.[5] Yang et al.[6] built per-expression PCA spaces and used them to exchange the expression of a face in a single image based on a different input image of the same subject.

The success of part-based PCA models suggests that localizing the surface region used for PCA results in improved detail and accuracy of the model. This led to the construction of a hierarchical space using wavelet transforms and applying PCA to the individual wavelet coefficients independently,[17] a technique first developed in medical imaging (see Section 2.3). A related pyramidal technique models high-frequency details such as wrinkles.[16]

Vlasic et al.[8] first used a *multilinear model* to represent 3D faces in multiple identities and multiple expressions. This model, based on tensor algebra, can be

6   *A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer*

thought of as a Cartesian product of PCA spaces–e.g. one for identity and one for expression, although higher-order models are possible. Dale et al.[9] used this model to compute a 4D sequence (3D+time) from video. This subsequently allowed to replace subjects in video performing the same expression. Yang et al.[10] used it to enhance or dampen expressions in videos. Bolkart and Wuhrer[11] learned such a model from a database of 3D scans of multiple subjects in different expression levels, and used it to register, both spatially and temporally, a database of 3D video sequences of subjects performing different expressions. The multilinear model was recently combined with the wavelet decomposition of the surface[18] to efficiently model fine-scale detail.

2.2. *Human Body Shapes*

Allen et al.[19] were the first to build a statistical model for 3D human body modeling. The shapes were acquired in a standard posture, and geometrically analyzed using a *global PCA model* over the vertex coordinates. This model can be applied to predict realistic 3D human body shapes in standard posture from sparse image information[20–22] or measurements.[23] Extensions of this basic global PCA model include part-based PCA models[35] and PCA models built on local shape representations instead of vertex coordinates.[34]

To jointly model shape and posture variation, Anguelov et al.[24] introduced the *SCAPE model*. This model combines a PCA space capturing shape variations learned from a database of subjects in standard posture with a mapping from posture parameters (based on joint angles) to changes in vertex coordinates learned from a database of one subject in multiple postures. The main idea behind SCAPE is to model body shape and posture as decorrelated. The SCAPE model has been successfully applied to various applications including the estimation of 3D human body shape and posture from (RGB or depth) images[25,26,29] and image and video modification.[27,28] Hirshberg et al.[30] jointly optimize a SCAPE model and point-to-point correspondences of a set of training data.

A different avenue of work models the correlation between shape and posture changes. This line of work is relevant, since the influence of the posture on the body shape depends on the subject, e.g. on the muscularity of the subject. Hasler et al.[31] proposed to jointly capture shape and posture variations by using PCA on a rotation-invariant encoding of a set of training shapes. A different way of analyzing shape and posture variations in correlation is to use a multilinear model.[33]

The reviewed global models for posture and shape variations have the disadvantage that areas near the joints may be distorted unless large databases for training are available that cover many shape and posture variations. To remedy this, part-based statistical models have been proposed. Chen et al.[36] relate the parts using a multilinear model, while Zuffi and Black[37] relate them using a graphical model.

2.3. *Medical Data*

In medical imaging, body parts, such as organs or parts thereof, are especially interesting shapes. One interesting task is to localize the shape of interest in a 2D or 3D medical image. This decomposition is typically called *segmentation* in this context. As for human faces and bodies, a global PCA model can help to solve this task. This was first explored by Cootes et al.,[38] and is called *active shape model*. This model has been applied successfully (e.g. Cootes and Taylor[39] and references therein). The registration of the shapes used for training has a strong influence on the quality of the active shape model. Davies et al.[40] jointly optimize the active shape model and the underlying registration using an information-theoretic approach based on description length. PCA uses a linear Euclidean subspace to represent the data. However, many data sets lie on a curved manifold rather than a Euclidean space. To find this manifold, Fletcher et al.[41] extend PCA to consider geodesic rather than Euclidean distances between shapes. This approach is called *principal geodesic analysis (PGA)*.

A problem of active shape models is their global support. In medical imaging, it is especially important to detect localized shape anomalies, as these can indicate disease, for instance. To this end, part-based models were proposed.[42,43] These models are applicable in problems where a meaningful segmentation into parts can be pre-defined. However, natural segmentations are difficult to define for many medical imaging tasks. To address this problem, statistical analysis using local wavelet representations has been used to learn a localized prior shape models and applied to segmentation problems in 2D and 3D images.[45–49]

3. Learning a Statistical Shape Space

To learn a statistical shape space, or to train a statistical shape model, we assume we are given T training shapes in full correspondence. In Sections 3.2 and 3.3 we describe the training process for two statistical shape spaces for human faces: global PCA and localized wavelet PCA. The key difference between them is the basis in which a prior probability distribution is fit to the training data.

We pre-align the data to remove rotation, translation, and uniform scale differences using generalized Procrustes analysis (GPA).[54] Note that by removing uniform scale differences, we only consider shape differences and not size differences of the models. For data-fitting this is desirable due to different measurement units used by different acquisition systems, but in general this is application dependent.[54] GPA iteratively aligns each model to the mean shape and recomputes the mean. Removing transformations that are not of interest using GPA is an important pre-processing step that yields better statistical models.

Learning a statistical shape space requires determining the basis functions or vectors, which define the shape space, and fitting a probability distribution to the resulting shape space coefficients from the training set. PCA-based methods per-

form these two steps simultaneously, selecting as a basis the directions of greatest variations in the data and computing a diagonal sample covariance matrix for the data projected onto this basis, which corresponds to a maximum likelihood estimate of a multi-dimensional Gaussian distribution. Part-based and wavelet-domain methods decompose the shapes into a localized basis before proceeding with PCA. The result is a basis consisting of the localized basis functions composed with the localized principal components. The learned prior is the product of the localized multi-dimensional Gaussian distributions. Note that the assumption of a Gaussian density function is only introduced when equipping the shape space with a Gaussian prior.

With this framework, and by restricting ourselves to linear shape spaces for the purposes of this study, the generator function Eq. (1), can be written as a combination of the basis functions

$$\mathbf{F}(\mathbf{s}) = \overline{\mathbf{F}} + \Phi\mathbf{s} = \overline{\mathbf{F}} + \sum_{i=1}^{d} \Phi_i s_i \tag{2}$$

where $\overline{\mathbf{F}}$ is the mean shape computed over the training set, $\Phi \in \mathbb{R}^{3n \times d}$ is a matrix, $\Phi_i \in \mathbb{R}^{3n}$ are its columns, and as before $\mathbf{s} \in \mathbb{R}^d$ is a vector of shape parameters. It is precisely the choice of Φ that determines the properties of the shape space, and determines the prior distribution learned from the training samples.

3.1. *Evaluating a Statistical Model*

So far, the number of basis functions d of the shape space is a free parameter. By choosing a small d, we will tend to lose shape detail, because the basis only spans a small portion of the variability of the training data. Choosing a large d, and keeping many basis functions, tends to result in a model that is overfitted to the training data. Overfitting is an issue when the model is underconstrained by the training samples, which may occur when the model has too many degrees of freedom, or when the model is learned from training samples with higher dimensionality than d.

To pick a number of basis functions d that preserves a high amount of variability yet does not overfit the training data, we use the following three error measures similar to *compactness*, *generalization*, and *specificity*.[55] We use a slight modification of the original error measures to obtain results that are independent of the size of the training data.

Compactness measures how much variability of the training data is explained by the learned statistical model. That is, we want to measure what fraction of the total variability of the training data is captured by d model parameters. This provides a measure of how well a given number of parameters explains the training data.

Generalization measures the ability of the model to represent data, which are not part of the training set. To calculate this measure, we learn a PCA model on a subset of the training data, where one subject is excluded. The excluded subject

is projected to the PCA space, reconstructed, and the distance between the source and the reconstruction is measured. To measure the distance between two faces, we use the average Euclidean vertex distance computed between all corresponding vertices. We perform this measurement for all subjects. The mean and standard deviation are then considered.

Specificity measures the similarity between reconstructions from the statistical model and the training data. This estimates the plausibility of a random face represented using the learned shape space. To calculate specificity we choose a set of random points sampled from the probability distribution of the learned statistical shape space. For each of these points we reconstruct the shape using Eq. (2) and compute the distance to the closest face in the training data. The distance between two faces is computed as above. The mean and standard deviation for the random sample are then considered.

3.2. *Global PCA*

Principal component analysis aims to reduce the complexity of a set of data. Due to its simplicity it is widely used for shape analysis. PCA is a linear transformation of a set of vectors from $\mathbb{R}^{3n}$ to $\mathbb{R}^d$ with $d < 3n$. A vector $\mathbf{f} \in \mathbb{R}^{3n}$ is expressed by the scalar weights s_i in a d-dimensional subspace, spanned by the orthogonal vectors $\mathbf{V}_i$, by

$$\mathbf{F}(\mathbf{s}) = \overline{\mathbf{F}} + \sum_{i=1}^{d} s_i \mathbf{V}_i. \tag{3}$$

For each parameterized shape of the training set we have one vector $\mathbf{F}_i^{(train)} \in \mathbb{R}^{3n}$ that contains an ordered coordinate set of all points of the i-th training shape. The vectors $\mathbf{V}_i$ are the eigenvectors of the data covariance matrix

$$\Sigma_{\mathbf{F}} = \frac{1}{T} \sum_{i=1}^{n} (\mathbf{F}_i^{(train)} - \overline{\mathbf{F}})(\mathbf{F}_i^{(train)} - \overline{\mathbf{F}})^T, \tag{4}$$

where $\overline{\mathbf{F}}$ is the mean of the training data. The eigenvectors $\mathbf{V}_i$ are ordered with respect to the non-increasing corresponding eigenvalues λ_i. The eigenvalues λ_i measure the variability captured by the i-th principal component. More specifically, $\mathbf{V}_i$ captures $100\frac{\lambda_i}{\sum_{i=1}^{T-1} \lambda_i}\%$ of the variability of the training data. The rank of the data covariance matrix is at most $\min(3n - 1, T - 1)$ and therefore the number of distinct non-zero eigenvalues and hence, the number or principal components, is at most $\min(3n - 1, T - 1)$.

Thus, we get our basis directly from the data via the principal components: in matrix form Φ_G has columns $\Phi_{Gi} = \mathbf{V}_i$, for $i = 1, \ldots, d$, where $d \leq \min(3n - 1, T - 1)$. Note that every basis vector Φ_{Gi} has global support in general: all $3n$ elements are in general non-zero, and every vertex is influenced.
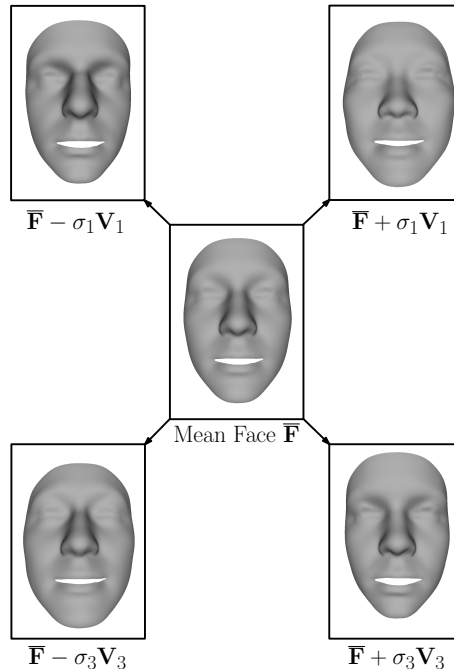
Fig. 1.   *Variations of two principal components.*

Properties Global PCA represents high-dimensional faces in a low-dimensional space, which is spanned by the eigenvectors corresponding to the largest d eigenvalues of the data covariance matrix. Figure 1 shows the variations along two principal components in the range of $-3\sigma_i$ to $+3\sigma_i$, where σ_i is the standard deviation of the i-th principal component.

Computing the statistical measures compactness, generalization and specificity for a global PCA model is straightforward. For example, compactness for d principal components is defined as

$$C\left(d\right) = \sum_{i=1}^{d} \lambda_i / \sum_{i=1}^{T-1} \lambda_i, \tag{5}$$

where λ_i is the i-th eigenvalue of the data covariance matrix.

3.3.   *Wavelet PCA*

Wavelet transforms project sampled data onto basis functions, which are localized in space and frequency. For statistical shape analysis, we can use a wavelet basis as a prefix, and extract a data driven basis for each individual wavelet coefficients using PCA.

In particular, we use second-generation or lifting wavelets[56] as they are highly efficient to compute. The input samples, the vertices of a face shape in our case,

are partitioned into maximally correlated subsets, and one subset is used to predict the other. The residual difference of an input sample from this prediction is called a detail, or wavelet, coefficient. The detail coefficients are then used to compute approximation, or scaling, coefficients from the other subset. The process is repeated hierarchically on the scaling coefficients.

Wavelets were originally defined on regularly sampled Euclidean domains,[57] but were extended to surfaces with regular subdivision sampling, known as spherical wavelets.[58] A popular such wavelet basis is a biorthogonal generalized B-spline basis,[59] which has been used in multiple statistical shape analysis applications.[17,47]

For more details about the forward and inverse transforms of the wavelet basis used here, we refer the reader to Bertram et al.[59] For more details on how they are used for statistical shape analysis, we refer the reader to Li et al.[47] and Brunton et al.[1,17]

Wavelet bases have an important property for statistical shape analysis as compared to other local bases: they decorrelate the data, which means we can compute statistics on each coefficient separately. Since the individual coefficients are of much lower dimension than the full surface, there is a much lower risk of overfitting for the same number of training samples T.

Performing PCA over the whole set of wavelet coefficients would result in the same principal components as the global model, because the wavelet transform is a linear transform, and PCA essentially just rotates the data so that the coordinate axes align with the directions of greatest variations. Instead, this method performs PCA locally on each coefficient, which is a 3D vector quantity, over the database.

First, let us denote the mean of each wavelet coefficient over the database by

$$\bar{\mathbf{s}}^k = \frac{1}{T} \sum_{i=1}^{T} \mathbf{s}_i^k, \tag{6}$$

where k indexes the coefficients.

While we can perform statistical analysis on each $\mathbf{s}^k$ independently of other values of k, we must consider their three components together. Each $\mathbf{s}^k$ is a 3D vector representing either the scale (absolute value) or the detail (relative value) of the shape at a particular frequency and spatial location. However, the coordinate axes in general do not correspond to the directions of greatest variation in the database. Therefore, we perform PCA on each set of coefficient vectors, to obtain 3D vectors $\mathbf{r}_i^k$ that represent the position along the directions of greatest variation, and 3×3 matrices U^k that transform these coordinates to our original world coordinate system, as in

$$\mathbf{s}_i^k = \bar{\mathbf{s}}^k + U^k \mathbf{r}_i^k \tag{7}$$

where we write $\mathbf{s}^k = [x_s^k, y_s^k, z_s^k]^T$ and $\mathbf{r}^k = [x_r^k, y_r^k, z_r^k]^T$ to denote the components of these vectors. Applying the transform $(U^k)^T$ to the data diagonalizes the covariance matrix, thus making each component independent.

The reconstruction of a face shape from the wavelet coefficients is performed via the inverse wavelet transform, which is a linear operator on the vector of concatenated wavelet coefficients $\mathbf{s}$, and can be written as a $3n \times 3n$ matrix D^{-1}. Thus, we can construct the combined basis Φ_W as

$$\mathbf{F}(\mathbf{s}) = D^{-1}\mathbf{s} \tag{8}$$

and from Eq. (7) we have

$$\mathbf{F}(\mathbf{s}) = D^{-1}\bar{\mathbf{s}} + D^{-1}U\mathbf{s} \tag{9}$$

where $\bar{\mathbf{s}}$ is the concatenation of the coefficient means $\bar{\mathbf{s}}^k$, and U is a block-diagonal $3n \times 3n$ matrix with the matrices U^k on the diagonal. Therefore, because $D^{-1}\bar{\mathbf{s}} = \overline{\mathbf{F}}$, we have

$$\Phi_W = D^{-1}U \tag{10}$$

as our combined basis, and the dimensionality of our shape space is $d = 3n$. Note that Φ_W has full rank.

Properties The local model has the benefit that it avoids overfitting, and as a consequence we can keep all variability present in the training set. Intuitively, the *local* surface properties of any given surface point are not likely to be specific to one set of faces or another. Whereas for the global model a bias in the training set, over-representation of one sex or a particular ethnicity or age range, can cause the lesser principal components to be highly specialized to that set, the geometry of a local surface patch is likely to be less dependent on the training data. The consequence is a somewhat unexpected behavior: by training and combining many low-dimensional models, which due to the limited flexibility of the training space ($\mathbb{R}^3$) have reduced sensitivity to bias in the training set, we get a final model with much greater flexibility, because truncation becomes unnecessary.

Figure 2 visualizes the mean shape color-coded with the magnitude of the shape variability for four levels of the wavelet subdivision, which corresponds to the localized shape variations at different scales. At finer scales, the variation quickly localizes around major facial features and reduces in magnitude.
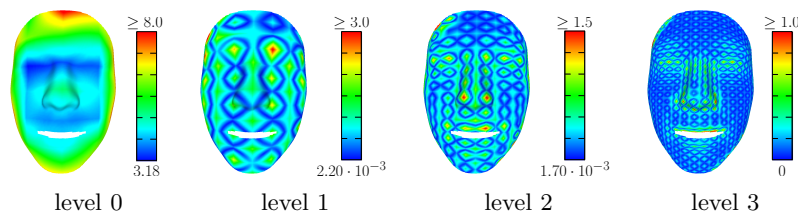

Fig. 2.   *Mean shape color-coded with the magnitude of shape variability for different levels. All units are in millimeters.*

The dimensionality of the local model is *statistically* more favorable. If, as is usually the case, the number of vertices is much greater than the number of

training examples, $n \gg T$, then the global model has problems of fitting to the particularities of the training set. In the local model, many independent statistical priors are learned, each with dimension 3. We have many more training examples than that. The independence of the local priors further allows an exhaustive search of the parameter space. Thus, we have no danger of getting trapped in local minima.

We can also consider the statistical measures from Section 3.1 for the wavelet PCA model. There is no simple formula for the compactness of this model, but since we retain all $3n$ shape parameters, and all variation, the compactness measure is 100%. Generalization and specificity can be computed the same way as for the global PCA model. Retaining all $3n$ parameters results in a generalization measure of 0.

4. Fitting a Statistical Shape Model

Once we have learned a statistical shape model, we can fit it to ambiguous data, such as noisy point clouds. The goal is to obtain a clean, registered shape, such as a face, from data that may be corrupted by noise, outliers, holes or occlusions. In this section, we present a generic model fitting approach framed as a constrained energy minimization over the learned model parameters.
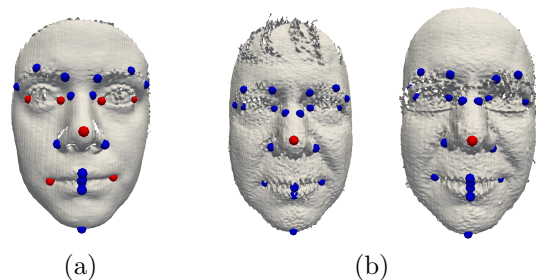

Fig. 3.   *Bosphorus scan with landmarks. (a) Red landmarks are used for initial alignment, blue landmarks are used for error evaluation. (b) Landmarks for two scans of the same identity. The position of the red landmark differs slightly for the two scans.*

4.1. *Initial Alignment*

To fit a statistical shape model to an input data set, we first need to align the input data and the statistical shape model to be in the same global coordinate system. Since we consider only shape differences in the training data, the initial alignment aims to find the rotation, translation, and uniform scaling that best aligns the statistical shape model with the input data.

To compute such an initial alignment, corresponding landmarks are commonly used. These landmarks can be manually located on the mean shape of the aligned training database once. On the input data, the landmarks can be predicted in a

fully automatic way.[60,61] However, since we use a test database that contains a set of landmarks, we choose to use a subset of these landmarks (the ones shown in red in Figure 3a) to compute an initial alignment. This approach removes a potential source of fitting error due to landmark prediction inaccuracies.

Another commonly used way to rigidly align two shapes is to use automatically detected features. We test a method of this flavor in our experiments. The method we use proceeds by finding corresponding features on the mean face and the input scan using spin images[62] and by performing random sample consensus.[63] This fully automatic method is expected to lead to less accurate alignments than the use of the given landmarks.

4.2. *Energy Minimization in Shape Space*

Our goal is to fit the statistical shape model to the input data as closely as possible while staying in the learned shape space. To fit our model to data, we minimize an energy function that amounts to the sum of squared distances between each model vertex and its nearest neighbor in the input point cloud. For our experiments, we use the following commonly used basic energy to pull the model towards the data

$$E_{data}(\mathbf{s}) = \sum_{i=1}^{n} \min \left(\left\| \mathbf{f}_i - \mathbf{p}_{\mathrm{NN}(i)} \right\|_2^2 , \tau \right) \qquad (11)$$

where $\mathbf{f}_i$ is vertex i of $\mathbf{F}(\mathbf{s})$ (see Eq. (1) and (2)), $\mathbf{p} \in P$ is a point in the input point cloud, $\mathrm{NN}(i)$ returns the index of the nearest neighbor in P of $\mathbf{f}_i$, and τ is a truncation threshold to add robustness against outliers. We compute nearest neighbors with a k-d tree using the implementation in ANN.[64]

When fitting a statistical shape model to data, the space of possible solutions should only contain likely shapes, thus ensuring that only plausible results are possible. A common and intuitive approach is to use the (negative logarithm of the) learned prior distribution as an energy term. In the case of PCA, it is common to assume a multi-dimensional Gaussian centered on the mean shape. In terms of energy minimization, this amounts to placing a soft constraint that the solution should be close to the mean. By design, however, this introduces a bias into the optimization, and results using this technique tend to lose distinctiveness and look similar to the mean.

Patel and Smith[65] proposed an alternative prior that is aimed at maintaining the distinctiveness of the models. They model the shape space as a manifold that is at a constant Mahalanobis distance from the mean. This is based on the observation that the squared Mahalanobis distances from the mean of a set of d-dimensional normally distributed vectors form a χ_d^2 distribution with expected value equal to d. Hence, Patel and Smith restrict the shapes to be on the hyper-ellipsoid at Mahalanobis distance $\sqrt{d}$ from the mean in order to preserve shape distinctiveness. While this approach models distinctiveness using the expected Mahalanobis distance from the mean, it does not consider the normal distributions along each dimension of the

shape space. That is, the modeled shape space may contain highly unlikely shapes along the directions of the principal components, as can be seen for the shape at the intersection of the hyper-ellipsoid shown in red and the x-axis in Figure 4.
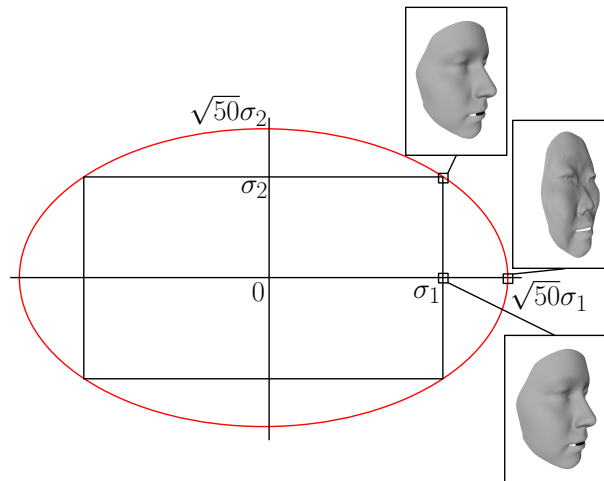

Fig. 4. *Unrealistic shapes may occur for large numbers of dimensions d along the directions of the principal components when maintaining a fixed Mahalanobis distance from the mean shape. Here, a global statistical shape space with $d = 50$ and $c = 1$ is shown.*

To avoid a bias towards the mean shape, while avoiding highly unlikely shapes, we simply place constraints on the shape parameters. That is, as a prior we use an indicator function on a region of the shape space. For PCA models, the ideal region would be an ellipsoid corresponding to an isosurface of a chosen minimum probability value. However, it is much easier and more efficient to impose linear constraints on an optimization. Therefore, we force the shape parameters to remain within the hyper-box of $\pm c\sigma_i$ about the mean, where σ_i is the standard deviation along axis i of the shape space, and c is a parameter specifying the amount shapes are allowed to deviate. Figure 4 shows a two-dimensional example. These constraints are equivalent to a prior probability of the form

$$P(\mathbf{s}) = \prod_{i=1}^{d} P_i(s_i) \tag{12}$$

where

$$P_i(s_i) = \begin{cases} 1 & |s_i| \le c\sigma_i \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

if we assume the shape parameters s_i are centered (mean subtracted). We call this a *hyper-box prior*. Combined with the data term in Eq. (11) it allows recovery of distinctive face shapes, while avoiding unrealistic ones.

A common post-process to fitting the parameters of the statistical models, is to then leave the statistical shape space and perform a fine-fitting of the vertex positions directly, similar to a template fitting method. We deliberately do *not* do this for two reasons. First, such a step is most often necessary when the learned shape space is not sufficiently generalizable to express novel shapes. This can occur due to insufficient or poor training data. However, as detailed in Section 5, we train from clean data containing a good sampling of both sexes and different ethnicities with a high-quality registration. Thus, our learned models are of high-quality.

Second, we wish to study the properties of the statistical shape spaces themselves, and leaving the space in a post-process would inevitably introduce additional uncertainty in analyzing the fitting results in Section 5. This is particularly true when we evaluate the fitting in the presence of occlusions.

5. Experiments

In this section, we evaluate the statistical models according to several criteria. For a more detailed evaluation, the reader is referred to Brunton et al.[1]

5.1. *Experimental Setup*

Training Data For training, we use the neutral expressions of $T = 100$ subjects from the BU-3DFE database.[66] This database contains relatively clean surfaces without occlusions, and a typical cropped face contains about 7500 vertices. Furthermore, each cropped face is equipped with 83 landmark points.

Parameterization We parametrize the database using the method of Salazar et al.[61] that deforms a template to each input face. This method is capable of predicting landmark points to aid in the template fitting. However, since we are given manually placed landmarks, our algorithm uses these instead of predicted ones. This removes a potential source of error during registration. The template we use contains 5996 vertices. We choose this low-resolution template for parametrization since the database has low resolution and does not contain small shape details. While the BU-3DFE database contains six additional expressions in four different levels, we consider only neutral expressions for our comparison. The resulting registration is of high-quality, which has been verified by manually inspecting each registered face.

Test Data We use a subset of 20 subjects (10 female and 10 male) of the Bosphorus database[67] to test the two models. Each subject is present in five occlusion levels: without occlusion, with glasses, with an occlusion of one eye by a hand, with an occlusion of the mouth by a hand, and with an occlusion of parts of the face by hair. Examples for each occlusion class can be seen in the left column of Figure 7. We chose this database as it allows the evaluation of different methods in the presence of severe occlusion. The resolution of this database is fairly high, and a typical face contains about 35000 vertices.

5.2. *Error Measures*

We evaluate the two statistical models using the error measures discussed in Section 3.1, by measuring the error when fitting to a surface via surface distance, and by visual evaluation.
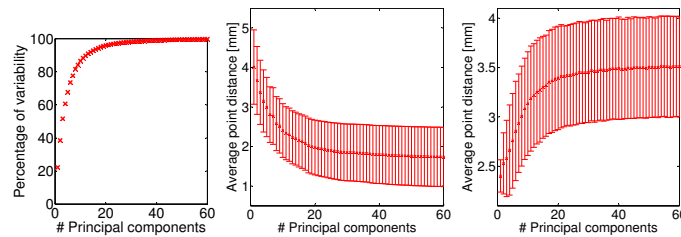

Fig. 5.   *Compactness, generalization and specificity for the global PCA model.*

Statistical Measures Ideally, we want a shape space that is compact, general and specific. However, generalization and specificity are often at odds with one another, and therefore we must find a balance by choosing the number of shape parameters d we retain. Figure 5 shows how the statistical error measures change as the number of principal components increases. We see that 30 principal components explain over 98% of the data variability in the training set. Further, for > 30 components, the decrease in generalization error is very slight, which suggests a limited benefit to select more components. Finally, for > 30 components, the specificity error still increases, which means the model represents plausible faces. We therefore choose dimensionality $d = 30$ for the global model. For the specificity error, we generate 10000 random samples and show the mean and standard deviation over all samples.

As mentioned in Section 3.3, the compactness and generalization of the wavelet model are fixed by the fact that we retain all variability in the model. We measured specificity in the same way as for the global model and obtained a mean point distance of 3.87mm with a standard deviation of 1.12mm. As the local model is less specific than the global model, this value is slightly higher than that of the global model.

Surface Distance To evaluate the fitting quality, we use the distance between each point in the fitted model and its closest point on the input scan. While this gives a lower bound on the fitting error of semantic correspondences between the model and the scan, it can be readily computed without a prior registration of the input.

Table 2 shows the surface distance statistics for the two models, and the influence of using different landmarks for evaluation–manually clicked landmarks available from the database, and automatically predicted landmarks using Spin image alignment. We see that the two sets of landmarks perform comparably.

We also see that the distributions of surface distances are different for the two models (Table 2 and Figure 6). The local model has lower mean and median errors, but higher standard deviations and maximum distances, as compared to the global

18   *A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer*

Table 2.   Surface distance statistics (in mm) for different initialization strategies.

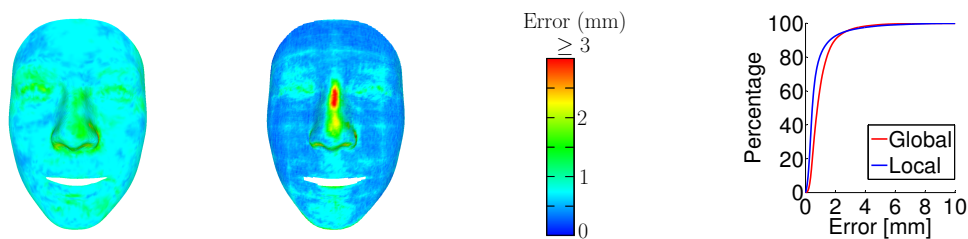| Initialization	Mean	Median	Std. Dev.	Max
Global Model				
manual landmarks	1.06	0.81	0.91	14.00
Spin image alignment	0.91	0.74	0.64	9.18
Local Model				
manual landmarks	0.80	0.47	1.06	15.63
Spin image alignment	0.63	0.43	0.72	16.51

Fig. 6.   *False color visualizations of the median nearest neighbor distances for different types of occlusion and cumulative error per vertex.*

model. The reflects the fact that the local model can fit very closely in some places, but if the initialization is too far away, the nearest neighbor threshold may prevent it from fitting closely to the surface in that location. Thus, we can conclude that while the local model allows recovery of more fine-scale detail, it is more sensitive to initialization.

Visual Qualitative Evaluation Finally, we evaluate the results visually. Figure 7 shows some examples of model fitting results – the top row without occlusion and the bottom row with an occlusion over the mouth. In each row, we see the input scan, the fitting result of the global model both smoothly shaded and color-coded with distance to the input data, and the fitting result of the local model in the same format. On the right is the color-coding legend. We see clearly the difference in behavior of the two models: the global model gets the overall shape, but omits fine details, while the local model gets a very good fit in most areas of the face, but in some locations is too far away after initialization to snap to the surface. For the occlusion case, we see how the occlusion of the mouth affects the shape of the whole face for the global model. In contrast, the local model localizes the influence of the occlusion.

6. Conclusion

In this chapter, we have given an overview of statistical shape spaces for 3D data, including a review of existing shape spaces, and a description of how they are learned and fit to raw 3D data. This is an important topic as more and more 3D
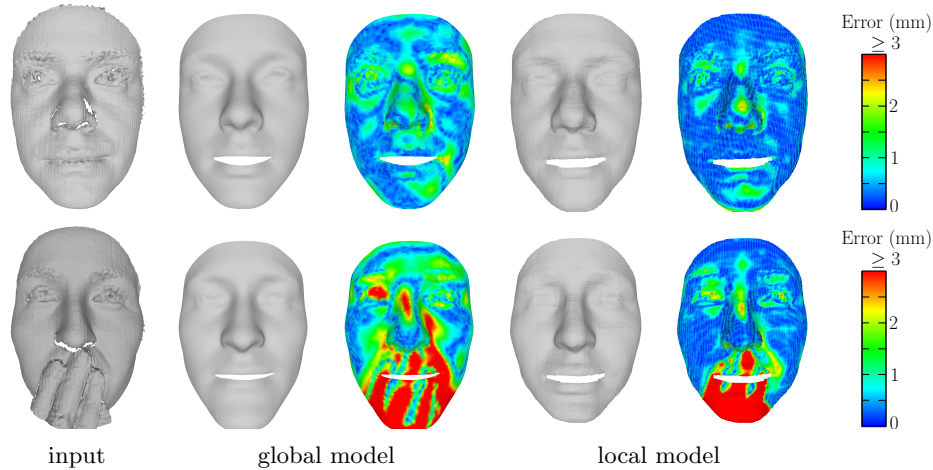
Fig. 7. *Some fitting results. Each row shows from left to right: input data, result of global fitting, color coding of distances between global fitting result and input data, result of local fitting, color coding of distances between local fitting result and input data, and the legend for the color coding.*

data becomes available, and scanning modalities become more affordable and widely used.

The main difference between statistical shape spaces is the basis in which shapes are represented. Almost universally, some form of PCA is used to obtain the main variations in the data. The difference comes in what transformation is applied to the shapes before PCA is applied. For object classes with multiple types, or modes, of shape variations, multilinear models are commonly used. While these models have already been studied in the literature, their exploration will no doubt increase in the future as dynamic 3D data becomes more widely available.

Correspondence or registration of a set of training shapes is a prerequisite for statistical shape analysis. A few methods jointly estimate shape and optimize a statistical model, thereby improving both. This is an increasingly popular research topic.

Acknowledgements

20   *A. Brunton, A. Salazar, T. Bolkart and S. Wuhrer*

References

1. A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer, Review of statistical shape spaces for 3D data with comparative analysis for human faces, *Computer Vision and Image Understanding.* **128**, (2014).
2. T. Bolkart, A. Brunton, A. Salazar, and S. Wuhrer. Statistical 3D shape models of human faces. http://statistical-face-models.mmci.uni-saarland.de/, (2013).
3. V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *ACM Conference on Computer Graphics and Interactive Techniques*, pp. 187–194, (1999).
4. B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, and T. Vetter. Reconstructing high quality face-surfaces using model based stereo. In *IEEE International Conference on Computer Vision*, pp. 1–8, (2007).
5. A. Patel and W. Smith. 3D morphable face models revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1327–1334, (2009).
6. F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas, Expression flow for 3D-aware face component transfer, *ACM Transactions on Graphics.* **30**(4), 60:1–10, (2011).
7. B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3D face recognition with a morphable model. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–6, (2008).
8. D. Vlasic, M. Brand, H. Pfister, and J. Popović, Face transfer with multilinear models, *ACM Transactions on Graphics.* **24**(3), 426–433, (2005).
9. K. Dale, K. Sunkavalli, M. Johnson, D. Vlasic, W. Matusik, and H. Pfister, Video face replacement, *ACM Transactions on Graphics.* **30**(6), 130:1–10, (2011).
10. F. Yang, L. Bourdev, J. Wang, E. Shechtman, and D. Metaxas. Facial expression editing in video using a temporally-smooth factorization. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 861–868, (2012).
11. T. Bolkart and S. Wuhrer, 3D faces in motion: Fully automatic registration and statistical analysis, *Computer Vision and Image Understanding.* **131**, 100–115, (2015).
12. C. Basso and A. Verri. Fitting 3D morphable models using implicit representations. In *Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 45–52, (2007).
13. F. ter Haar and R. Veltkamp. 3D face model fitting for recognition. In *European Conference on Computer Vision*, pp. 652–664, (2008).
14. M. Smet and L. V. Gool. Optimal regions for linear model-based 3D face reconstruction. In *Asian Conference on Computer Vision*, pp. 276–289, (2010).
15. I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karamelpatzis, and T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **29**(4), 640–649, (2007).
16. A. Golovinskiy, W. Matusik, H. Pfister, S. Rusinkiewicz, and T. Funkhouser, A statistical model for synthesis of detailed facial geometry, *ACM Transactions on Graphics.* **25**(3), 1025–1034, (2006).
17. A. Brunton, C. Shu, J. Lang, and E. Dubois. Wavelet model-based stereo for fast, robust face reconstruction. In *Canadian Conference on Computer and Robot Vision*, pp. 347–354, (2011).
18. A. Brunton, T. Bolkart, and S. Wuhrer. Multilinear wavelets: A statistical shape space for human faces. In *European Conference on Computer Vision*, (2014).
19. B. Allen, B. Curless, and Z. Popović, The space of human body shapes: reconstruction and parameterization from range scans, *ACM Transactions on Graphics.* **22**(3), 587–

594, (2003).

20. H. Seo, Y. I. Yeo, and K. Wohn, 3D body reconstruction from photos based on range scan, *Technologies for E-Learning and Digital Entertainment.* pp. 849–860, (2006).

21. Y. Chen and R. Cipolla. Learning shape priors for single view reconstruction. In *IEEE International Workshop on 3-D Digital Imaging and Modeling*, pp. 1425–1432, (2009).

22. J. Boisvert, C. Shu, S. Wuhrer, and P. Xi, Three-dimensional human shape inference from silhouettes: Reconstruction and validation, *Machine Vision and Applications.* **24**(1), 145–157, (2013).

23. S. Wuhrer and C. Shu, Estimating 3D human shapes from measurements, *Machine Vision and Applications.* **24**(6), 1133–1147, (2013).

24. D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, Scape: shape completion and animation of people, *ACM Transactions on Graphics.* **24**(3), 408–416, (2005).

25. P. Guan, A. Weiss, A. O. Balan, and M. J. Black. Estimating human shape and pose from a single image. In *International Conference on Computer Vision*, pp. 1381–1388, (2009).

26. A. Balan and M. Black. The naked truth: Estimating body shape under clothing. In *European Conference on Computer Vision*, pp. 15–29, (2008).

27. S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, Parametric reshaping of human bodies in images, *ACM Transactions on Graphics.* **29**(4), 126:1–10, (2010).

28. A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt, MovieReshape: Tracking and reshaping of humans in videos, *ACM Transactions on Graphics.* **29**(5), 148:1–10, (2010).

29. A. Weiss, D. Hirshberg, and M. Black. Home 3D body scans from noisy image and range data. In *International Conference on Computer Vision*, pp. 1951–1958, (2011).

30. D. Hirshberg, M. Loper, E. Rachlin, and M. Black. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In *European Conference on Computer Vision*, pp. 242–255, (2012).

31. N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, A statistical model of human pose and body shape, *Computer Graphics Forum.* **28**(2), 337–346, (2009).

32. N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, and H.-P. Seidel, Estimating body shape of dressed humans, *Computers & Graphics.* **33**(3), 211–216, (2009).

33. N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Multi-linear pose and body shape estimation of dressed subjects from image sets. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1823–1830, (2010).

34. S. Wuhrer, C. Shu, and P. Xi, Posture-invariant statistical shape analysis using laplace operator, *Computers & Graphics.* **36**, 410–416, (2012).

35. P. Xi, W.-S. Lee, and C. Shu. A data-driven approach to human body cloning using a segmented body database. In *Pacific Graphics*, pp. 139–147, (2007).

36. Y. Chen, Z. Liu, and Z. Zhang. Tensor-based human body modeling. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 105–112 (June, 2013).

37. S. Zuffi and M. Black. The stitched puppet: A graphical model of 3D human shape and pose. In *IEEE International Conference on Computer Vision and Pattern Recognition*, (2015).

38. T. Cootes, C. Taylor, D. Cooper, and J. Graham, Active shape models – their training and application, *Computer Vision and Image Understanding.* **61**(1), 38–59, (1995).

39. T. Cootes and C. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *SPIE Medical Imaging*, pp. 236–248, (2001).

40. R. Davies, C. Twining, and C. Taylor, *Statistical Models of Shape: Optimisation and*

Evaluation. (Springer, 2008).

41. P. Fletcher, C. Lu, S. Pizer, and S. Joshi, Principal geodesic analysis for the study of nonlinear statistics of shape, *IEEE Transactions on Medical Imaging.* **23**(8), 995–1005, (2004).

42. M. Toews, D. Collins, and T. Arbel. A statistical parts-based appearance model of inter-subject variability. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 232–240, (2006).

43. F. Lecron, J. Boisvert, S. Mahmoudi, H. Labelle, and M. Benjelloun. Fast 3D spine reconstruction of postoperative patients using a multilevel statistical model. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 446–453, (2012).

44. C. Davatzikos, X. Tao, and D. Shen, Hierarchical active shape models, using the wavelet transform, *IEEE Transactions on Medical Imaging.* **22**(3), 414–423, (2003).

45. D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3D shape analysis using spherical wavelets. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 459–467, (2005).

46. D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Shape-driven 3D segmentation using spherical wavelets. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 66–74, (2006).

47. Y. Li, T.-S. Tan, I. Volkau, and W. Nowinski. Model-guided segmentation of 3D neuroradiological image using statistical surface wavelet model. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–7, (2007).

48. S. Essafi and G. Langs. Hierarchical 3D diffusion wavelet shape priors. In *IEEE International Conference on Computer Vision*, pp. 1717–1724, (2009).

49. P. Yu, B. T. T. Yeo, P. E. Grant, B. Fischl, and P. Golland. Cortical folding development study based on over-complete spherical wavelets. In *IEEE International Conference on Computer Vision*, pp. 1–8, (2007).

50. O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, A survey on shape correspondence, *Computer Graphics Forum.* **30**(6), 1681–1707, (2011).

51. G. Tam, Z.-Q. Cheng, Y.-K. Lai, F. Langbein, Y. Liu, D. Marshall, R. Martin, X.-F. Sun, and P. Rosin, Registration of 3D point clouds and meshes: A survey from rigid to non-rigid, *IEEE Transactions on Visualization and Computer Graphics.* **19**(7), 1199–1217, (2013).

52. T. Cashman and A. Fitzgibbon, What shape are dolphins? Building 3D morphable models from 2d images, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **35**, 232–244, (2013).

53. D. A. Alcantara, O. Carmichael, W. Harcourt-Smith, K. Sterner, S. R. Frost, R. Dutton, P. Thompson, E. Delson, and N. Amenta, Exploration of shape variation using localized components analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **31**(8), 1510–1516, (2009). ISSN 0162-8828.

54. I. Dryden and K. Mardia, *Statistical Shape Analysis.* (Wiley, 2002).

55. M. Styner, K. Rajamani, L.-P. Nolte, G. Zsemlye, G. Szekely, C. Taylor, and R. Davies. Evaluation of 3D correspondence methods for model building. In *Information Processing in Medical Imaging*, pp. 63–75, (2003).

56. W. Sweldens, The lifting scheme: A custom-design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis.* **3**(2), 186–200, (1996).

57. S. Mallat, *A Wavelet Tour of Signal Processing.* (Elsevier, 1999).

58. P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *ACM Conference on Computer Graphics and Interactive Techniques*, pp. 161–172, (1995).

59. M. Bertram, M. Duchaineau, B. Hamann, and K. I. Joy, Generalized B-Spline subdivision-surface wavelets for geometry compression, *IEEE Transactions on Visualization and Graphics.* **10**(3), 326–338, (2004).

60. C. Creusot, N. Pears, and J. Austin. 3D face landmark labelling. In *ACM Workshop on 3D Object Retrieval*, pp. 27–32, (2010).

61. A. Salazar, S. Wuhrer, S. Chu, and F. Prieto, Fully automatic expression-invariant face correspondence, *Machine Vision and Applications.* **25**(4), 859–879, (2014).

62. A. Johnson and M. Hebert. Recognizing objects by matching oriented points. In *Conference on Computer Vision and Pattern Recognition*, pp. 684–692, (1997).

63. M. Fischler and R. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM.* **24**(6), 381–395, (1981).

64. D. Mount and S. Arya. ANN: A library for approximate nearest neighbor searching. http://www.cs.umd.edu/ mount/ANN/, (2010).

65. A. Patel and W. Smith. Exploring the identity manifold: Constrained operations in face space. In *European Conference on Computer Vision*, pp. 112–125, (2010).

66. L. Yin, X. We, Y. Sun, J. Wang, and M. Rosato. A 3D facial expression database for facial behavior research. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 211–216, (2006).

67. A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *Workshop on Biometrics and Identity Management*, pp. 47–56, (2008).