Approximating Intersections and Differences Between Linear Statistical Shape Models Using Markov Chain Monte Carlo

Maximilian Weiherer, Finn Klein, Bernhard Egger
Department of Computer Science
Friedrich-Alexander-Universität Erlangen-Nürnberg
maximilian.weiherer@fau.de

Abstract

To date, the comparison of Statistical Shape Models (SSMs) is often solely performance-based, carried out by means of simplistic metrics such as compactness, generalization, or specificity. Any similarities or differences between the actual shape spaces can neither be visualized nor quantified. In this paper, we present a new method to qualitatively compare two linear SSMs in dense correspondence by computing approximate intersection spaces and set-theoretic differences between the (hyper-ellipsoidal) allowable shape domains spanned by the models. To this end, we approximate the distribution of shapes lying in the intersection space using Markov chain Monte Carlo and subsequently apply Principal Component Analysis (PCA) to the posterior samples, eventually yielding a new SSM of the intersection space. We estimate differences between linear SSMs in a similar manner; here, however, the resulting spaces are no longer convex and we do not apply PCA but instead use the posterior samples for visualization. We showcase the proposed algorithm qualitatively by computing and analyzing intersection spaces and differences between publicly available face models, focusing on gender-specific male and female as well as identity and expression models. Our quantitative evaluation based on SSMs built from synthetic and real-world data sets provides detailed evidence that the introduced method is able to recover ground-truth intersection spaces and differences accurately.

1. Introduction

Statistical Shape Models (SSMs) are a popular class of generative models providing a low-dimensional parametric representation of complex objects. SSMs and especially 3D Morphable Models (3DMMs) [2] are widely used within the computer vision community and often applied to model humans (faces, bodies, bones, and organs). Their application ranges from face recognition [3], single-shot 3D face [14, 25, 40] and body reconstruction [29], face reenactment [37] and visual dubbing, to applications in the medical domain [17, 31, 39], forensics, cognitive science, neuroscience, and psychology [12]. SSMs are typically built by applying Principal Component Analysis (PCA) to a set of objects in dense correspondence. As such, an SSM is a *linear* model in which shapes are represented as points in a low-dimensional, affine vector space. Although other, non- or multi-linear models exist, PCA-based SSMs are still most common. They are easy to interpret, convenient to visualize, have excellent extrapolation capabilities, and are compatible with standard computer graphics pipelines. Hence, we consider only PCA-based models in this work. In general, however, the presented method can be used with all models that (i) use an affine vector space as an approximation for the underlying (manifold) shape space, and (ii) allow random sampling from and projecting shapes onto that space, including PCA-based Point Distribution Models [8] and Gaussian Process Morphable Models [31], multi-linear [4] and wavelet-based models [6], non-linear models based on Principal Geodesic Analysis (by applying our method on the tangent space) [16], but also more modern approaches based on Variational Autoencoders [24].

To date, comparison of SSMs is typically solely performance-based and carried out by applying metrics such as compactness, generalization, and specificity [10, 36]. A qualitative and direct comparison of the actual shape spaces spanned by two models is currently not possible; a visualization can only be provided for the individual shape spaces by inspecting random samples or the first principal modes of variation. As such, any differences or similarities between two models' shape spaces can not be computed nor visualized with existing metrics.

In this paper, we present a new approach to qualitatively compare two linear SSMs that takes into account the affine vector spaces spanned by the models. Specifically, we aim at computing the *intersection* and *set-theoretic difference* between the models' (hyper-ellipsoidal) allowable shape domains, *i.e.*, those shapes that can either be explained by both SSMs or only by one model but not the other. Given the extreme difference in dimensions between the low-dimensional shape spaces and their high-dimensional embedding, we propose to use a sampling-based technique to approximate the desired spaces. Starting with a geometric motivation and formal definition, we formulate a probability distribution over shapes lying in the intersection or set-theoretic difference and use approximate inference based on Markov chain Monte Carlo (MCMC) to generate samples from those spaces. Based on these samples, we compute the mean and a basis for the intersection space using PCA, eventually forming a new SSM that can be used for visualization, data exploration, and analysis. Since the set-theoretic difference of two models' subspaces is no longer convex, we do not apply PCA but only visualize posterior samples. Finally, note that thanks to the probabilistic nature of our method, we are not only able to inspect differences and similarities on shapes as a whole but also on a *per-vertex* level, allowing us to understand which *features* are present in both models and/or what is unique.

Our method has several interesting applications beyond a general data exploration use case. For instance, differences between race-specific models (*e.g.*, between Asian and White models) could allow for conclusions about demographic bias in SSMs. In the medical domain, differences and similarities between two models built from a healthy and pathological group will help to visualize and identify novel phenotypes or shape-based clinical indicators. Moreover, the difference between those two models or gender-specific shape variations may be of great interest as it could be later added to other models to increase variability, thus acting as a data augmentation strategy. Besides medical applications, we also see opportunities in product design. When modeling body parts, the comparison of two SSMs from distinct populations may be helpful to improve the design for a specific market. Finally, the difference between face identity and expression models would also reduce the effects of the identity-expression ambiguity in 3DMMs of faces [13], allowing for better expression neutralization or transfer.

To summarize, the key contributions of this paper are: (i) we present a new method to qualitatively compare two SSMs by computing approximate intersection spaces and set-theoretic differences between the low-dimensional allowable shape domains spanned by the models, (ii) we provide an extensive, quantitative evaluation using models built from synthetic data sets and SSMs which are publicly available, (iii) we analyze intersections and differences between popular face models, including gender-specific male and female and identity and expression models, and (iv) we show how our method can be used to approximate intersection spaces and differences between texture models of 3DMMs.

2. Related Work

Although tackling a different problem, most related to our work is the method proposed by Hall et al. [22]. They present a *splitting operation* used to remove one linear subspace from another and argue that this operation is seen as the inverse of the union operation. It is important to understand that this method yields again a linear and convex subspace; they do *not* compute the *set-theoretic difference* between two linear models' shape spaces. Rather, they split one subspace from the other, effectively keeping the intersection in that space. In contrast, the set-theoretic difference we are interested in excludes the intersection from this space (this is why we end up with a non-convex space). This allows us to investigate the *true* difference between two models, *i.e.*, those shapes that are exclusively represented in one model, which is not possible with [22].

Performance-based metrics. The most well-known performance-based metrics to compare SSMs are compactness, generalization, and specificity [10, 36]. Those metrics allow to measure how well a model represents a certain population; they enable quantitative comparison of two models of the same population and are applied in most of the publications presenting novel SSMs. However, although considered state-of-the-art when comparing SSMs, they do not allow for a visual inspection of similarities or differences between actual shape spaces. Babalola et al. [1] proposed to use the Bhattacharya distance to measure the overlap between the two distributions implied by SSMs. Although this might be useful in practice, it merely returns a scalar-valued distance. Another common way to compare models is by evaluating their performance on downstream applications [5, 18, 26, 35]. This, however, does not tell anything about what makes one model different or superior to another from a probabilistic point of view. In general, none of the existing approaches enable visualization of similarities or differences between low-dimensional shape spaces.

Computing intersections between vector spaces. Linear algebra offers an analytic way to compute the intersection between two subspaces embedded in a common vector space. Given bases $A, B \in \mathbb{R}^{m \times n}$ for two subspaces, a basis of the intersection space can be calculated by computing the null space of $C = [A, -B] \in \mathbb{R}^{m \times 2n}$. When considering SSMs, however, we usually have $m \gg n$, *i.e.*, there is an extreme difference in dimensions between the spanned subspaces of the models and their embedding. In this case, the linear system that we would need to solve when computing the null space of C is highly over-determined and may not have a non-trivial solution in general.

A lot of classical mathematical theory [21, 23] is available to compute the orthogonal *projection* matrix, P, used to project points onto the intersection of two subspaces. Numerous formulas and methods have been proposed to compute P, of which the alternating projection method [38] is probably most well known. However, this method has a high computational complexity and gives only the projection matrix instead of a basis for the intersection space.

Another method to calculate a basis for the intersection between two subspaces is by means of the Zassenhaus algorithm [30]. More recently, Fenggang et al. [15] proposed a closed-form algorithm to compute such a basis.

3. Method

After fixing notations and reviewing some basic properties of linear shape models, we first provide a formal definition of the intersection between two SSMs. We then present the algorithm to estimate the intersection. Additionally, we show how the same strategy can be further explored to estimate the differences between two models.

3.1. Preliminaries

A linear shape model can be interpreted either by assuming a linear algebraic or a probabilistic point of view. Firstly, we take the linear-algebraic approach and represent an SSM as a *function* of the form $f : \mathbb{R}^q \longrightarrow \mathbb{R}^{dn}$, $f(\alpha) = \bar{x} + U\alpha$, where $\bar{x} \in \mathbb{R}^{dn}$ is the so-called *mean shape* computed over a set of objects described with n points in d dimensions (for triangular meshes, $d = 3$). The orthogonal matrix $U = (\sqrt{\lambda_1}u_1, \sqrt{\lambda_2}u_2, \ldots, \sqrt{\lambda_q}u_q) \in \mathbb{R}^{dn \times q}$ holds the scaled eigenvectors of the dataset's covariance matrix, where λ_i is the i-th eigenvalue. The vector space spanned by a linear model is the affine subspace

$$M := \bar{x} + \mathrm{span}(U) = \{\bar{x} + U\alpha \mid \alpha \in \mathbb{R}^q\} \subseteq \mathbb{R}^{dn}, \quad (1)$$

where $\dim(M) = q$. Every linear combination of the basis vectors contained in U gives rise to a new shape.

We now describe the probabilistic perspective on linear shape models. By assuming $\alpha \sim \mathcal{N}(0, I)$, one can verify that shapes $x = f(\alpha)$ are distributed according to $x \sim \mathcal{N}(\bar{x}, C)$, where $C = UU^T \in \mathbb{R}^{dn \times dn}$ is the sample covariance matrix. The probability of a shape is given by

$$p(x) = p(\alpha) \propto \exp\left(-\frac{1}{2}\|\alpha\|_2^2\right). \quad (2)$$

From this perspective, not every shape in M is equally likely; the probability mass is concentrated in a q-dimensional hyper-ellipsoid which is centered at the mean $\bar{x}$ and whose principal axes correspond to the eigenvectors of C. The eigenvalues of C are the reciprocals of the squares of the lengths of the semi-axes. The shapes lying inside the hyper-ellipsoid can be characterized as $Q := \{x \in M \mid (x - \bar{x})^T C^{-1}(x - \bar{x}) \leq k\} \subseteq \mathbb{R}^{dn}$ with $k \in [0, 1]$. In terms of a probabilistic interpretation, this set can be equivalently rewritten as

$$Q = \{x \in M \mid p(x) \geq \xi\}, \quad (3)$$

where $\xi \in [0, 1]$ (see supp. material for a derivation). It contains all shapes with a probability greater than a certain threshold. Following [7] and for a suitable ξ, shapes in Q are considered *plausible* in the sense that they look similar to the observed (training) data; hence, Q is often called the *allowable shape domain* [8].

3.2. Computing Intersections

Given two aligned[1] SSMs in dense correspondence, our goal is to compute their intersection. We define the intersection $I \subseteq \Omega = M_1 \cup M_2$ between two SSMs as the intersection between their allowable shape domains, *i.e.*,

$$I := Q_1 \cap Q_2. \quad (4)$$

In this work, we consider only SSMs with non-empty I (whether two models intersect can be checked using, *e.g.*, [19]). As seen from (4) and by noting that $I = (M_1 \cap M_2) \cap (Q_1 \cap Q_2)$, a shape $x \in \Omega$ belongs to the intersection if the following conditions are met: **First**, x can be *represented* by both models, *i.e.*, $x \in M_1 \cap M_2$. Assuming $x \in M_1 \subset \Omega$, this is equivalent to finding an $x' \in M_2$ such that the (Euclidean) distance $d(x', x)$ vanishes (or vice versa if we assume $x \in M_2$). **Second**, x is *likely* in both models, that is, $p(x) \geq \xi_1$ and $p(x') \geq \xi_2$.

Our definition is motivated by the fact that most of the shapes in the subspaces M_1 and M_2 are unlikely and do not lead to realistic shape instances. Hence, if we defined the intersection simply as $I = M_1 \cap M_2$, then I would contain a lot of degenerated shapes that we want to disregard. We therefore only consider the plausible regions of two models, which are exactly identified by Q_1 and Q_2.

The problem with the definition in (4), however, is that real-world models are usually *not* noise-free. As a consequence, the first condition, $x \in M_1 \cap M_2$, will never be met in practice since we can not find an $x' \in M_2$ such that $d(x', x) = 0$ exactly. To account for this, we weaken the strong definition and formulate an *approximate* intersection I_ϵ for $\epsilon > 0$ as

$$I_\epsilon := (Q_1 \cap (Q_2 + B_\epsilon(0))) \cup (Q_2 \cap (Q_1 + B_\epsilon(0)))\,, \quad (5)$$

where $B_\epsilon(0) = \{y \in \mathbb{R}^{dn} : d(0, y) < \epsilon\}$ is the ϵ-ball in $\mathbb{R}^{dn}$ centered at 0. This allows us to also consider shapes to be in the intersection that are *almost* (up to ϵ) contained in I. We can rewrite (5) as

$$\begin{aligned} I_\epsilon = &\{x \in Q_1 \mid \exists x' \in Q_2 : d(x', x) < \epsilon\} \\ &\cup \{x \in Q_2 \mid \exists x' \in Q_1 : d(x', x) < \epsilon\}\,. \end{aligned} \quad (6)$$

Of course, $I_\epsilon \to I$ if $\epsilon \to 0$ and $I \subset I_\epsilon$ for every $\epsilon > 0$.

[1] By *aligned*, we mean that Euclidean similarity transformations (rotation, translation, and scaling) between the two models have been removed. This can always be achieved by aligning the respective mean shapes using ordinary Procrustes analysis, see, *e.g.*, [11].

3.3. Algorithm

Instead of explicitly constructing the set I_ϵ, we aim at estimating the *distribution* of points $x \in I_\epsilon$. Given such distribution, we can then effectively generate samples from the intersection, which we use to build an SSM for the intersection by applying PCA. The resulting linear model can then be utilized to visualize and study the intersection space.

The advantage of estimating the distribution of $x \in I_\epsilon$ instead of randomly sampling from Ω and applying hard constraints to test whether a point x belongs to the intersection is that we do *not* need to explicitly set values for the parameters (ϵ, ξ_1, and ξ_2) involved in (6).

In detail, we model the posterior distribution of all $x \in Q_1$, given that $x \in I_\epsilon$ (and vice versa for all $x \in Q_2$):

$$p(x \,|\, x \in I_\epsilon) = \frac{p(x \in I_\epsilon \,|\, x) p(x)}{\int p(x \in I_\epsilon \,|\, x) p(x) \mathrm{d}x}. \tag{7}$$

We call $p(x \in I_\epsilon \,|\, x)$ the likelihood function and denote it by $L(x; x \in I_\epsilon)$. The probability $p(x)$ is computed according to (2) by noting that $x = f_1(\alpha)$ for $\alpha \in \mathbb{R}^{q_1}$.

The likelihood function encodes the conditions under which x belongs to the intersection space. In the following, we show how the two conditions as stated in Section 3.2 can be directly translated into a likelihood.

First condition. We implement the first condition, *i.e.*, finding an $x' \in M_2$ such that $d(x', x)$ becomes small, using the orthogonal projection of x onto M_2, given by

$$x' = \mathrm{proj}_{M_2}(x) = f_2(\underbrace{U_2^{-1}(x - \bar{x}_2)}_{=\alpha' \in \mathbb{R}^{q_2}}). \tag{8}$$

The point x' is the closest point to x contained in M_2 and hence minimizes the distance between x and x'. With that, we define a distance likelihood L_D as

$$L_D(x; x \in I_\epsilon) \propto \exp\left(-\frac{1}{2}\left(\frac{d(x', x)}{\sigma}\right)^2\right), \tag{9}$$

where we used

$$d(x', x) = \frac{1}{n}\sum_{i=1}^{n} \|x_i' - x_i\|_2 \tag{10}$$

and assumed $d(x', x) \sim \mathcal{N}(0, \sigma^2)$. A similar likelihood was originally proposed in [34] and recently used by [32] for model-based surface registration.

Second condition. For the second condition, we require both x and x' to be likely. This can be implemented by taking into account the probabilities $p(x)$ and $p(x')$ of x and x'. Since the probability of x is already included in the formulation of the posterior in (7), we only need to care about x'. To enforce the projection x' to be likely, too, we define a projection likelihood L_P simply as

$$L_P(x; x \in I_\epsilon) = p(x'), \tag{11}$$

where the probability of $x' = f_2(\alpha')$ is given by (2). Our final likelihood is then a combination of the distance and projection likelihood and formulated as

$$L = L_D L_P. \tag{12}$$

The posterior distribution $p(x \,|\, x \in I_\epsilon)$ in (7) is now fully specified. However, it is intractable and can not be computed exactly. We therefore make use of approximate inference as described in the following.

3.4. Implementation

We use Markov chain Monte Carlo (MCMC) to approximate the intractable posterior distribution in (7). This is possible since $p(x \,|\, x \in I_\epsilon) \propto L(x; x \in I_\epsilon) p(x)$.

To construct a Markov chain, we make use of the Metroplis-Hastings (MH) algorithm. The MH algorithm requires a proposal distribution conditioned on the current state. We use the same random walk mixture proposal as proposed in [34]. Please refer to the supp. material for more information about the MH algorithm and our proposal distribution. To explore the parameter space and to reduce auto-correlation between samples as much as possible, we use an ensemble of Markov chains instead of only one chain. Each chain has a different starting point which is sampled from $\mathcal{N}(0, I)$. We ensured that samples are sufficiently far apart to avoid individual chains exploring the same part of the parameter space.

3.5. Computing Differences

Keeping in mind the definition of the approximate intersection space, we define the set-theoretic difference between two SSMs as

$$D_{12} \coloneqq Q_1 \setminus (M_2 + B_\epsilon(0)). \tag{13}$$

The set D_{12} contains all $x \in Q_1$ that can *not* be represented in M_2 (or vice versa for D_{21}). We can rewrite (13) into

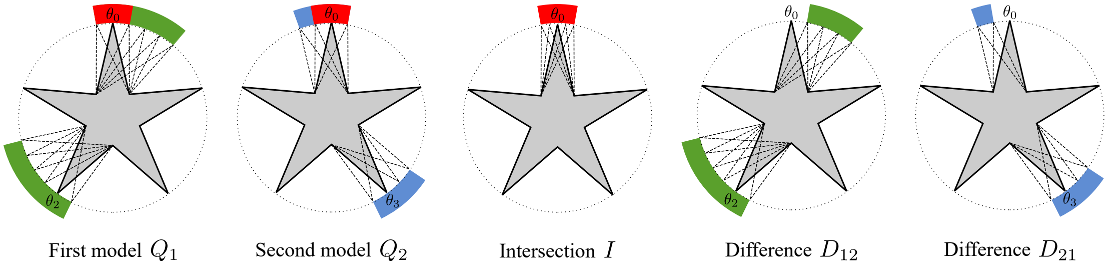
$$D_{12} = \{x \in Q_1 \mid \forall x^* \in M_2 : d(x^*, x) \geq \epsilon\}. \tag{14}$$

From a computational perspective, this formulation is rather unpleasant due to the universal quantifier involved. Note, however, that the point $x' = \mathrm{proj}_{M_2}(x)$ minimizes the distance from x to M_2 and thus, if $d(x', x) \geq \epsilon$, then the same holds for all $x^* \in M_2$. With this, we can rephrase (14) as

$$D_{12} = \{x \in Q_1 \mid d(x', x) \geq \epsilon\}. \tag{15}$$

The condition of whether x belongs to the difference can now be easily checked by means of the projection operator.

To estimate the distribution of points lying in D_{12} (or D_{21}), we only have to make two small changes to the algorithm presented throughout the previous sections. From the definition of the difference in (15) we observe that the



First model Q_1 Second model Q_2 Intersection I Difference D_{12} Difference D_{21}

Figure 1. Schematic illustration of how we generate training data for star models including ground-truth intersections and differences. For the first model, we vary the first and third point of the star in $[\theta_j - a, \theta_j + b]$, where $j \in \{0, 2\}$. For the second model, we vary the first and the fourth point in $[\theta_j - c, \theta_j + a]$, where $j \in \{0, 3\}$. The ground-truth intersection contains stars where the first point varies from $[\theta_0 - a, \theta_0 + a]$, whereas the ground-truth difference of the first and the second model contains stars where the first point ranges from $(\theta_0 + a, \theta_0 + b]$ and the third between $[\theta_2 - a, \theta_2 + b]$. The ground-truth difference of the second and the first model includes stars where the first point varies from $[\theta_0 - c, \theta_0 - a)$ and the fourth from $[\theta_3 - c, \theta_3 + a]$.

only condition a point $x \in Q_1$ has to fulfill in order to lie in D_{12} is that the distance to its projection onto M_2 becomes large. This can be implemented by *inverting* the distance likelihood in (9) as

$$\bar{L}_D(x; x \in I_\epsilon) \propto 1 - L_D(x; x \in I_\epsilon). \qquad (16)$$

Secondly, since we no longer require the projection to be likely, we remove the projection likelihood from (12).

After those minor modifications, our algorithm is ready to be used to approximate the distribution of shapes lying in the difference between two models. However, the space in which the shapes of the difference between two SSMs lie is no longer convex as the shapes lying in the intersection have been removed. As such, it does not make any sense to apply PCA to the MCMC samples as the resulting linearized space would simply interpolate the intersection. Instead, we only visualize random samples drawn from our Markov chain in order to inspect the difference.

4. Experiments and Results

We conducted several quantitative and qualitative experiments to validate our method.

4.1. Quantitative Analysis

Our quantitative evaluation is based on two data sets from which we generate ground-truth intersection spaces. The first is a synthetic data set based on a five-pointed star. Due to the rather simplistic geometry, this data set is well suited for an initial analysis and serves as a proof-of-concept. The second is a real data set derived from a real-world SSM and, consequently, shares a lot of desired properties such as high dimensionality.

We employ the Grassmann distance as a natural distance between two affine subspaces to compare ground-truth intersections with the spaces recovered by our method. The Grassmann distance is defined as the Euclidean norm of the affine principal angles between two subspaces and can be easily computed by applying Singular Value Decomposition to the Stiefel coordinates of the two spaces, see, *e.g.* [28]. Please refer to the supp. material for more information.

Constructing ground-truth differences is far more complex due to their topological structure. As such, the quality of the differences recovered by our method can only be evaluated on the star models since, due to their synthetic construction, we are able to generate samples from the ground-truth difference (*i.e.*, samples for which we definitely know they are lying in the difference). This is not the case for real-world SSMs because *the ground-truth difference space for real-world models is unknown.*

4.1.1 Synthetic Data Set

For the synthetic data set, we use a five-pointed star as a simple geometric object to generate training data by systematically varying its points, see Figure 1. Each point of the star can be written as $(r\cos\theta_i, r\sin\theta_i)$ with radius r and angle θ_i, $i = 0, \ldots, 4$. For the first model, we vary the first and third point in range $[\theta_j - a, \theta_j + b]$, where $j \in \{0, 2\}$. For the second model, we vary the first and fourth point in range $[\theta_j - c, \theta_j + a]$, where $j \in \{0, 3\}$. The ground-truth intersection model contains stars where the first point ranges from $[\theta_0 - a, \theta_0 + a]$. The ground-truth difference of the first and the second model contains stars where the first point ranges from $(\theta_0 + a, \theta_0 + b]$ and the third between $[\theta_2 - a, \theta_2 + b]$. The ground-truth difference of the second and the first model includes stars where the first point varies from $[\theta_0 - c, \theta_0 - a)$ and the fourth from $[\theta_3 - c, \theta_3 + a]$.

We tested our method on six different values for a, b, c as shown in Table 1. An ensemble of 15 chains was used, where each chain was sampled 2,500 times. 1,000 samples were considered as burn-in. Out of all samples, 5,000 samples were evenly chosen to build an SSM for the intersection space. For the distance likelihood, we set $\sigma = 0.003$ for all

Training data					**Intersection** (Grassmann distances, $d_G \downarrow$)				**Difference** (Reconstruction errors, $d_R \uparrow$)			
Q_1	Q_2	I	D_{12}	D_{21}	$d_G(\hat{I}, I)$	$d_G(Q_1, I)$	$d_G(Q_2, I)$	$d_G(Q_1, Q_2)$	$d_R(\hat{D}_{12}, I)$	$d_R(\hat{D}_{21}, I)$	$d_R(D_{12}, I)$	$d_R(D_{21}, I)$
$[-5, 40]$	$[-20, 5]$	$[-5, 5]$	$(5, 40]$	$[-20, -5)$	0.0260 ± 0.0086	1.5634	1.5634	2.1211	0.2235 ± 0.1132	0.1148 ± 0.0534	0.1781 ± 0.2243	0.0911 ± 0.1122
$[-5, 20]$	$[-20, 5]$	$[-5, 5]$	$(5, 20]$	$[-20, -5)$	0.0370 ± 0.0110	1.5583	1.5583	2.1211	0.1165 ± 0.0556	0.1171 ± 0.0542	0.0911 ± 0.1122	0.0911 ± 0.1122
$[-10, 40]$	$[-20, 10]$	$[-10, 10]$	$(10, 40]$	$[-20, -10)$	0.0140 ± 0.0052	1.5471	1.5506	2.1211	0.2081 ± 0.1097	0.1150 ± 0.0540	0.1867 ± 0.2110	0.1073 ± 0.1081
$[-10, 20]$	$[-20, 10]$	$[-10, 10]$	$(10, 20]$	$[-20, -10)$	0.0323 ± 0.0094	1.5506	1.5506	2.1211	0.1107 ± 0.0510	0.1143 ± 0.0534	0.1126 ± 0.1080	0.1073 ± 0.1081
$[-20, 60]$	$[-30, 20]$	$[-20, 20]$	$(20, 60]$	$[-30, -20)$	0.0053 ± 0.0022	1.5240	1.5440	2.1211	0.2692 ± 0.1571	0.1509 ± 0.0809	0.2834 ± 0.3170	0.1861 ± 0.1569
$[-40, 80]$	$[-50, 40]$	$[-40, 40]$	$(40, 80]$	$[-50, -40)$	0.0018 ± 0.0004	1.4863	1.5375	2.1211	0.3257 ± 0.1994	0.2243 ± 0.1293	0.4200 ± 0.3993	0.3501 ± 0.2531

Table 1. Quantitative results for the star models based on different training sets, averaged over five runs. The comparison of the estimated intersection space, $\hat{I}$, and the ground-truth intersection I is based on the Grassmann distance, denoted as d_G. We also report various baseline distances to give an intuition for the range of the distances. To evaluate whether or not posterior samples from the computed differences, $\hat{D}_{12}$ and $\hat{D}_{21}$, are indeed from the true difference, we calculate reconstruction errors d_R by projecting estimated samples and samples from the ground-truth differences D_{12} and D_{21} into the ground-truth intersection model I. Note that the intervals for the training data arise from setting different values for the parameters a, b, and c.

experiments (see supp. material for an ablation on σ).

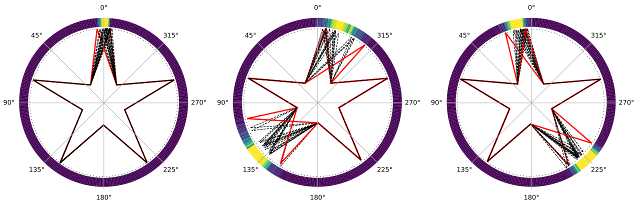
We use the Grassmann distance as described above to measure how well our method can recover the ground-truth intersection space. To evaluate the quality of the computed differences, we exploit the fact that we have *samples* from the ground-truth difference model (clearly, we do *not* have a parametrization of these models, but only a few samples thereof). To this end, we quantify whether or not samples from the difference generated by our method have as high a distance from their reconstruction in the ground-truth intersection space as the samples from the ground-truth difference (note that a sample belongs to the difference if it can *not* be represented in the intersection). We, therefore, project the MCMC samples of the difference into the ground-truth intersection model and calculate the average reconstruction error, see supp. material for further details. We do the same for the samples from the ground-truth difference. Similar reconstruction errors then indicate that all samples generated by our method are indeed valid samples from the true difference space.

Overall, the proposed method is able to recover the ground-truth intersection space for all six models very well, see Table 1. We measure an average Grassmann distance of 0.0194 between ground-truth intersections and the intersection spaces recovered by our method. We observe a similar result for the estimated differences in terms of reconstruction errors. Figure 2 provides additional visualization of the distribution of posterior samples.

4.1.2 Real Data Set

The simple geometry and ground-truth intersection spaces of the star models enabled an intuitive analysis of the proposed method. The geometry of real-world SSMs, however, is oftentimes far more complex. We conduct experiments on those models to further validate the proposed approach.

Given an SSM with a set of basis vectors, we generate two individual models and a ground-truth intersection by randomly splitting the set of basis vectors into three disjoint subsets, S_1, S_2, and S_I. The first two sets, S_1 and S_2, contain the unique basis vectors for the first and second model, and the last set, S_I, holds the basis vectors of the ground-truth intersection. The two models' bases are then given by $S_1 \cup I$ and $S_2 \cup I$, respectively. Since we can not generate samples from the ground-truth difference, we present evaluations only for the computation of intersection spaces.



Intersection $\hat{I}$ Difference $\hat{D}_{12}$ Difference $\hat{D}_{21}$

Figure 2. Visualization of (a random subset of) MCMC samples generated to compute the estimated intersection $\hat{I}$ and differences $\hat{D}_{12}$ and $\hat{D}_{21}$ between two star models. The included polar frequency histogram visualizes the distribution of posterior samples; it indicates that most of the samples are indeed from the true posterior distribution (*i.e.*, they are within the bounds shown in red).

Training data			**Intersection** (Grassmann distances, $d_G \downarrow$)			
$\dim(Q_1)$	$\dim(Q_2)$	$\dim(I)$	$d_G(\hat{I}, I)$	$d_G(Q_1, I)$	$d_G(Q_2, I)$	$d_G(Q_1, Q_2)$
3	3	1	0.0032 ± 0.0010	1.5517	1.5532	2.1854
6	6	2	0.0645 ± 0.0455	2.1641	2.1855	3.0835
9	9	3	0.0877 ± 0.0178	2.6375	2.6901	3.7747
12	12	4	0.1683 ± 0.0106	3.1004	3.0849	4.3563
15	15	5	0.1939 ± 0.0433	3.4606	3.4467	4.8682

Table 2. Quantitative results for the estimated intersection spaces of the real-world data set. We tested five different configurations of individual model dimensions, $\dim(Q_1)$ and $\dim(Q_2)$, and the dimension of the ground-truth intersection space, $\dim(I)$. Results were averaged over three different random splits per configuration.

We used the first 15 basis vectors of the BFM 2019 [18] for this experiment. Moreover, an ensemble of 25 chains was used. Each chain was sampled 5,000 times with a burn-in phase of 2,000 samples. To build the intersection model, again 5,000 samples were evenly chosen out of all samples. We empirically set $\sigma = 0.3$ for all models.

Table 2 presents the results for five different random

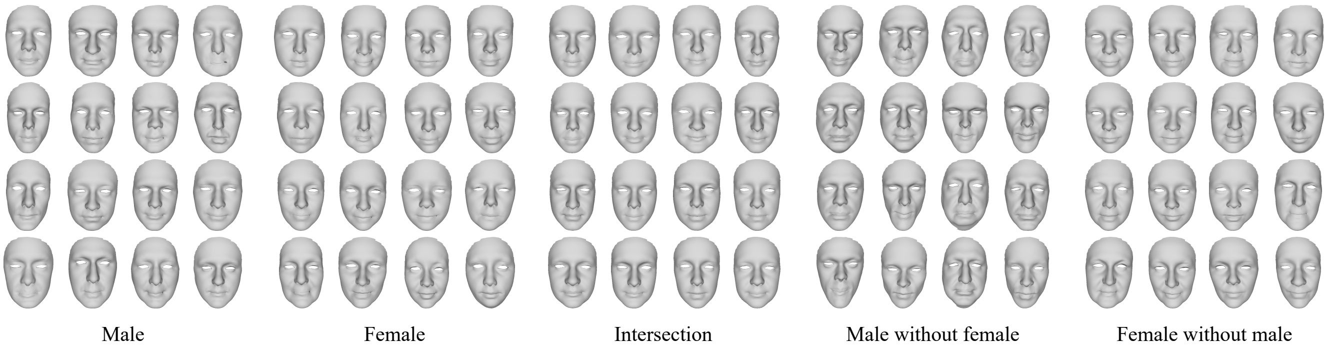


Figure 3. Random samples from the male and female model of the LYHM data set (1st and 2nd block) as well as samples drawn from the computed intersection model (3rd block) and the respective differences (male without female, and female without male, 4th and 5th block). We observe stronger male and female dominance in the differences and neutral gender in the intersection.

splits. As shown, the proposed method is able to recover all ground-truth intersection spaces quite well with an average Grassmann distance of 0.1035. However, we observe increasing distances as the dimension of the intersection increases. This is most likely because MCMC does not properly sample the entire space, and may be alleviated by choosing larger proposals or running more Markov chains. Indeed, by doubling the number of chains, the Grassmann distance decreases by more than 10%.

4.2. Qualitative Results

To showcase our method on real-world SSMs, we compute and analyze intersections and differences between (i) a gender-specific male and female model, and (ii) an identity and expression model. Due to the missing ground-truth intersection spaces and differences, we can only provide qualitative results. However, to underline that our method does *not* simply produce the union of two models, or even worse, just reproduces one model, we do report Grassmann distances to the union and to individual models. The union of two models is computed by applying PCA to a set of random samples from both models.

Male and female models. Both models were built using 600 faces from the LYHM database [9]. We used a face mask to only include the frontal region of the face and truncated both models to include the first 50 basis vectors. Each model has 5,764 vertices. As such, both SSMs span a 50-dimensional affine subspace which is embedded in a 17,292-dimensional vector space.

Exemplary results can be found in Figure 3. As seen, random faces sampled from the intersection model do not show a strong preference towards male or female, and most of the time tend to look more neutral compared to faces drawn from the individual male or female models. Furthermore, while random faces from the difference between male and female show extremely masculine traits, samples from the difference between female and male look rather feminine. We report a Grassmann distance of about 1.8 from the intersection to the union of male and female, and a distance of 3.2 and 2.8 to individual male and female subspaces. Computing the intersection model took about 110 minutes on a single core; differences half the time.

Identity and expression models. Face identity and expression are usually modeled separately in 3DMMs. As recently demonstrated in [13], however, the subspaces spanned by the identity and expression model are far from being orthogonal and thus, face identities and expressions are not independent of each other. As a consequence, it is not possible to alter a specific face identity without leaving its expression *completely* unchanged (and vice versa). Although only hardly visible with the human eye, this effect has implications on downstream applications such as expression transfer and inverse rendering [13].

We use our method to further study this ambiguity, also known as the identity-expression ambiguity, and analyze the intersection as well as differences between the identity and expression models of the BFM 2019 [18]. We again truncated both SSMs to the first 50 basis vectors. Each model has 1,746 vertices, leading to a 5,238-dimensional vector space in which the models' subspaces are embedded.

Results are shown in Figures 4 and 5. While there is no ground-truth for this task, the observations match our intuition, *e.g.*, shape variations of the nose do not arise in the expression model but solely in the identity model, and effects based on muscle movement dominate in the identity model. To still add a quantitative element, we measured the Grassmann distances between the intersection and original identity and expression models and report a distance of 5.36 to identity, and 5.09 to the BFM expression model. Those distances being close to each other show that the intersection model somehow lies between the identity and expression model. The Grassmann distance to the union is 3.48. It took about 30 minutes to compute the intersection model. Again, computing differences required half the time.

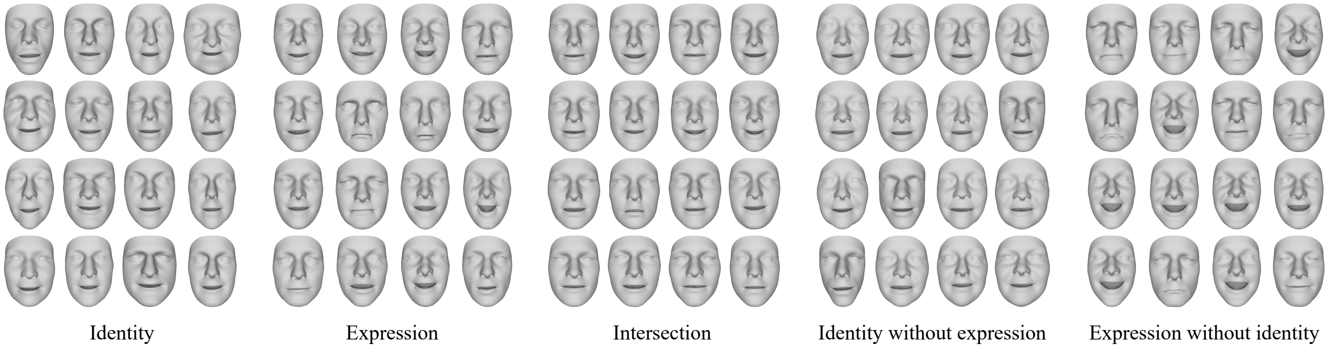


Figure 4. Random samples from the original identity and expression models of the BFM 2019 (1st and 2nd block) as well as samples of the computed intersection model (3rd block). Also shown are samples from the difference between identity and expression models (4th and 5th block). We observe less expression variation in samples drawn from the difference of identity and expression (identity without expression, 4th block) and less identity variation in the difference of expression and identity (expression without identity, 5th block).

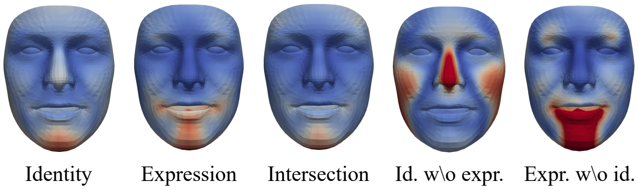


Figure 5. Per-vertex level visualization of similarities and differences between the BFM 2019 [18] identity and expression models. Depicted is the per-vertex variance of the original identity and expression models as well as the posterior variance for the computed intersection model and differences (red areas show high variance, blue areas low). The intersection visualizes the source of the identity-expression ambiguity, differences highlight *features* represented either by identity (nose) or expression (opening of mouth, raising cheeks and eyebrows).

For both tasks (male vs. female and identity vs. expression), we observed very similar results for the FLAME [27] model. Please refer to the supp. material.

4.3. Extension to Color

Since texture in 3DMMs is modeled similarly to shape [2], our method can be used to also compute intersections and differences between texture models *without* any algorithmic changes. Results for male and female 3DMMs described above can be found in the supp. material.

5. Limitations

Although the presented method yields appealing results, our approach in its current form is not without limitations. First, with increasing dimension of the intersection space more Markov chains are needed to explore the parameter space; this phenomenon is known as the curse of dimensionality. It naturally increases the run time of our method. We plan to tackle this issue by parallelizing Markov chains and by using more effective MCMC samplers such as the Metropolis-adjusted Langevin algorithm [20].

Second, if SSMs are not in dense correspondence, we require some form of optimization in the projection operation, *i.e.*, (8) needs to be adapted. Clearly, while the projection can be computed in closed-form if models are in correspondence, we require registration if they are not in correspondence. This can be *efficiently* carried out *prior* to the application of our method using techniques proposed in [33].

Lastly, although our method and the chosen parameters might not be trivial to evaluate on real-world models, the result can *always* be well validated empirically by estimating ϵ from the MCMC samples, see supp. material.

6. Conclusion

In this paper, we have introduced a new method to compare SSMs by computing approximate intersection spaces and set-theoretic differences between the low-dimensional allowable shape domains spanned by two models in dense correspondence. We showed how MCMC can be leveraged to approximate the distribution of shapes lying inside those spaces, and, based on posterior samples, compute a new SSM for the intersection. Building upon this algorithm, we further demonstrated that our method can be easily adapted to also compute set-theoretic differences. Confirmed by quantitative and qualitative evaluation on synthetic data and real-world face models, the proposed method is able to recover ground-truth intersection spaces and differences accurately. Moreover, qualitative results obtained on real-world SSMs match our intuition. In future work, we plan to study how our approach is (or may be) correlated to performance-based metrics, such as the Bhattacharya distance.

Acknowledgments. We thank Tinashe Mutsvangwa, Marcel Lüthi, and Tobias Meier for interesting discussions and feedback. This work was funded by the German Federal Ministry of Education and Research (BMBF), FKZ: 01IS22082 (IRRW), and the FAU Emerging Talents Initiative (ETI). The authors are responsible for the content of this publication.

References

[1] Kolawole O Babalola, Timothy F Cootes, Brian Patenaude, Anil Rao, and Mark Jenkinson. Comparing the similarity of statistical shape models using the bhattacharya metric. In *MICCAI*, pages 142–150, 2006. 2

[2] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, pages 187–194, 1999. 1, 8

[3] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *IEEE TPAMI*, 25(9):1063–1074, 2003. 1

[4] Timo Bolkart and Stefanie Wuhrer. A robust multilinear model learning framework for 3d faces. In *CVPR*, pages 4911–4919, 2016. 1

[5] James Booth, Anastasios Roussos, Stefanos Zafeiriou, Allan Ponniah, and David Dunaway. A 3d morphable model learnt from 10,000 faces. In *CVPR*, pages 5543–5552, 2016. 2

[6] Alan Brunton, Timo Bolkart, and Stefanie Wuhrer. Multilinear wavelets: A statistical shape space for human faces. In *ECCV*, pages 297–312, 2014. 1

[7] Timothy F Cootes and Cristopher J Taylor. Statistical models of appearance for computer vision, 2004. Technical report, University of Manchester. 3

[8] Timothy F Cootes, Cristopher J Taylor, David H Cooper, and Jim Graham. Active shape models – their training and application. *Comput. Vis. Image. Underst.*, 61(1):38–59, 1995. 1, 3

[9] Hang Dai, Nick Pears, William AP Smith, and Christian Duncan. Statistical modeling of craniofacial shape and texture. *IJCV*, 28(2):547–571, 2019. 7

[10] Rhodri H Davies. Learning shape: optimal models for analysing natural variability, 2002. PhD Thesis, University of Manchester. 1, 2

[11] Ian L Dryden and Kanti V Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 2016. 3

[12] Bernhard Egger, William AP Smith, Ayush Tewari, Stefanie Wuhrer, Michael Zollhöfer, Thabo Beeler, Florian Bernard, Timo Bolkart, Adam Kortylewski, Sami Romdhani, Christian Theobalt, Volker Blanz, and Thomas Vetter. 3d morphable face models - past, present and future. *ACM TOG*, 39(5):1–38, 2020. 1

[13] Bernhard Egger, Skylar Sutherland, Safa C Medin, and Joshua Tenenbaum. Identity-expression ambiguity in 3d morphable face models. In *FG*, pages 1–7, 2021. 2, 7

[14] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM TOG*, 40(4):1–13, 2021. 1

[15] Yan Fenggang, Liu Shuai, Wang Jun, and Jin Ming. Closed-form algorithms for computing the intersection of two subspaces. *J. Syst. Eng. Electron.*, 30(2):245–250, 2019. 3

[16] P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging*, 2004. 1

[17] Jean-Rassaire Fouefack, Bhushan Borotikar, Tania S Douglas, Valérie Burdin, and Tinashe EM Mutsvangwa. Dynamic multi-object gaussian process models. In *MICCAI*, pages 755–764, 2020. 1

[18] Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. Morphable face models-an open framework. In *FG*, pages 75–82, 2018. 2, 6, 7, 8

[19] Igor Gilitschenski and Uwe D Hanebeck. A robust computational test for overlap of two arbitrary-dimensional ellipsoids in fault-detection of kalman filters. In *Inf. Fusion*, pages 4911–4919, 2012. 3

[20] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *J. R. Stat. Soc. Ser. B Methodol.*, 73(2):123–214, 2011. 8

[21] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU Press, 2013. 2

[22] Peter Hall, David Marshall, and Ralph Martin. Merging and splitting eigenspace models. *IEEE TPAMI*, 22(9):1024–1049, 2000. 2

[23] William N Anderson Jr and Richard J Duffin. Series and parallel addition of matrices. *J. Math. Anal. Appl.*, 26(3):576–594, 1969. 2

[24] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014. 1

[25] Chunlu Li, Andreas Morel-Forster, Thomas Vetter, Bernhard Egger, and Adam Kortylewski. Robust model-based face reconstruction through weakly-supervised outlier segmentation. In *CVPR*, pages 372–381, 2023. 1

[26] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM TOG*, 36(6):1–17, 2017. 2

[27] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. In *ACM TOG*, pages 194:1–194:17, 2017. 8

[28] Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye. The grassmannian of affine subspace. *Found. Comput. Math.*, 21(2):537–574, 2020. 5

[29] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM TOG*, 34(6):1–16, 2015. 1

[30] Eugene M Luks, Ferenc Rákóczi, and Charles RB Wright. Some algorithms for nilpotent permutation groups. *J. Symb. Comput.*, 23(4):335–354, 1997. 3

[31] Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter. Gaussian process morphable models. *IEEE TPAMI*, 40(8):1860–1873, 2017. 1

[32] Dennis Madsen, Andreas Morel-Forster, Patrick Kahr, Dana Rahbania, Thomas Vetter, and Marcel Lüthi. A closest point proposal for mcmc-based probabilistic surface registration. In *ECCV*, pages 281–296, 2020. 4

[33] Stylianos Ploumpis, Haoyang Wang, Nick Pears, William AP Smith, and Stefanos Zafeiriou. Combining 3d morphable models: A large scale face-and-head model. In *CVPR*, pages 10934–10943, 2019. 8

[34] Sandro Schönborn, Bernhard Egger, Andreas Morel-Forster, and Thomas Vetter. Markov chain monte carlo for automated face image analysis. *IJCV*, 123(2):160–183, 2017. 4

[35] William AP Smith, Alassane Seck, Hannah Dee, Bernard Tiddeman, Joshua B Tenenbaum, and Bernhard Egger. A morphable face albedo model. In *CVPR*, pages 5011–5020, 2020. 2

[36] Martin Styner, Kumar Rajamani, Lutz-Peter Nolte, Gabriel Zsemlye, Gábor Székely, Christopher Taylor, and Rhodri Davies. Evaluation of 3d correspondence methods for model building. In *IPMI*, pages 63–75, 2003. 1, 2

[37] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *CVPR*, pages 2387–2395, 2015. 1

[38] John von Neumann. On rings of operators. reduction theory. *Ann. Math.*, 50(2):401–485, 1949. 2

[39] Maximilian Weiherer, Andreas Eigenberger, Bernhard Egger, Vanessa Brébant, Lukas Prantl, and Christoph Palm. Learning the shape of female breasts: an open-access 3d statistical shape model of the female breast built from 110 breast scans. *The Vis. Comput.*, 39:1597–1616, 2023. 1

[40] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Towards metrical reconstruction of human faces. In *ECCV*, pages 250–269, 2022. 1