Deformation Models for Human Shape Analysis

Stefanie Wuhrer

Deformation Models for Human Shape Analysis

Stefanie Wuhrer

Synthèse des travaux scientifiques pour obtenir le grade de

Habilitation à Diriger des Recherches.

Summary

This manuscript presents the author's most significant works conducted from 2011 to 2017 on the topic of processing and analyzing 3D geometric data, and in particular deformations of 3D human shapes and their accessories. Applications of such 3D data include product design, avatar creation, and recognition tasks. Inferring high-level information from raw scan data is challenging due to high levels of noise in the captured data and a high degree of variability in geometry both across different human subjects and across different poses. For this reason, to date, most of the 3D scan data used in applications are processed with the help of manual input. The goal of our work can be summarized as providing automatic methods for processing and analyzing raw geometric 3D data showing the shape and deformations of humans and their accessories.

To achieve this goal, we take advantage of the fact that humans and their clothing allow for a limited set of deformations, and use deformation models to constrain the space of possible human 3D shapes. This document presents contributions based on three deformation models. The first part considers a near-isometric deformation model defined on partial regions that can be used to model loco-motions of humans as well as cloth deformations. The main novelty of this model is that it allows for processing that is robust to acquisition noise. The second and third parts consider generative models for human body and face shape that are learned from a training database of 3D scans. For both bodies and faces, we present automatic processing pipelines that allow to build generative models based on databases of thousands of raw scans. We further present applications of both the near-isometric deformation model and the generative model of human body shape.

Résumé

Ce manuscrit présente les travaux les plus significatifs de l'auteur réalisés de 2011 à 2017 sur le traitement et l'analyse des données géométriques 3D, et en particulier les déformations des formes humaines 3D et de leurs accessoires. Les applications de ces données 3D comprennent la conception de produits, la création d'avatars et les tâches de reconnaissance. Il est difficile d'inférer des informations de haut niveau à partir de données scannées brutes en raison des niveaux élevés de bruit dans les données capturées et d'un degré élevé de variabilité géométrique entre différents sujets humains et entre différentes poses. Pour cette raison, à ce jour, la plupart des données numérises en 3D utilisées dans les applications sont traitées à l'aide d'une saisie manuelle. Le but de notre travail peut être résumé comme fournissant des méthodes automatiques pour le traitement et l'analyse des données 3D géométriques brutes montrant la forme et les déformations des humains et de leurs accessoires.

Pour atteindre cet objectif, nous profitons du fait que les humains et leurs vêtements permettent un ensemble limité de déformations, et utilisons des modèles de déformation pour contraindre l'espace des formes 3D humaines possibles. Ce document présente des contributions basées sur trois modèles de déformation. La première partie considère un modèle de déformation quasi-isométrique défini sur des régions partielles qui peuvent être utilisées pour modéliser des locomotions humaines ainsi que des déformations de tissu. La nouveauté principale de ce modèle est qu'il permet un traitement robuste au bruit d'acquisition. Les deuxième et troisième parties considèrent des modèles génératifs pour le corps humain et la forme du visage qui sont appris à partir d'une base de données. Pour les corps et les visages, nous présentons des pipelines de traitement automatique qui permettent de construire des modèles génératifs à partir de bases de données de milliers de scans 3D bruts. Nous présentons en outre des applications du modèle de déformation quasi-isométrique et du modèle génératif de la forme du corps humain.

Acknowledgments

The work presented in this manuscript was achieved in collaboration with excellent colleagues, and I would like to thank all of them for our great collaborations. I would like to especially thank my (former) Ph.D. students Victoria Fernández Abrevaya, Timo Bolkart, Abdullah Haroon Rasheed, Aurela Shehu and Jinlong Yang.

Contents

1 Introduction

Problem statement and motivation

This work is concerned with processing and analyzing 3D geometric data. In recent years, measured data such as 3D scans are increasingly used to build and analyze high-quality 3D models in applications ranging from entertainment in computer graphics to ergonomic design and medical applications. One reason is that it has recently become increasingly affordable to digitize 3D geometric information, while the quality of such acquisitions has greatly increased. Today, 3D data can be acquired with the help of e.g. image-based reconstruction systems, 3D laser-range scanners, or structured light systems (see e.g. [32] for a review of scanning techniques), and it is now possible to digitize moving shapes along with their appearance information at high spatial and temporal resolution (see e.g. [35, 70]). However, the resulting acquisitions are unstructured and require further processing and analysis to extract high-level information from the data.

A particularly interesting class of shapes are human models along with their accessories. Applications of such 3D shapes include product design, avatar creation, and recognition tasks among others. Inferring high-level information from raw scan data is challenging in this scenario due to high levels of noise in the captured data and the high variability in geometry both across different human subjects and across different poses as illustrated in Figure 1.1. For this reason, to date, most of the 3D scan data used in applications is processed with the help of manual input given e.g. in the form of sparse annotations of the data.

We aim to allow for an automatic processing and analysis of raw geometric data showing the shape and deformations of humans and their accessories. To achieve this goal, we take advantage of the fact that humans and their clothing allow for a limited set of deformations, which can be used to constrain

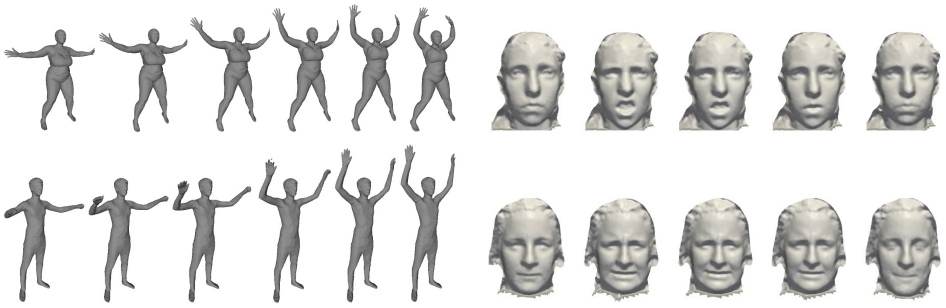

Figure 1.1: *Recent acquisition technologies allow to capture 3D human body (left) and face (right) shape and motion. However, automatically processing and analyzing such data is challenging. (Data on top left from [90], on bottom left captured using the Kinovis platform at Inria Grenoble Rhône-Alpes, and on right from [139].)*

the space of possible 3D human shapes. This in turn allows to reduce the search space of subsequent processing tasks. In this work, we consider three distinct deformation models to this end.

Structure and overview

The most general deformation model that is commonly used in 3D shape analysis is to require the deformation between two shapes $\mathcal{S}_1$ and $\mathcal{S}_2$ to be a *homeomorphism*, which is the case if the mapping $f : \mathcal{S}_1 \to \mathcal{S}_2$ describing the deformation from $\mathcal{S}_1$ to $\mathcal{S}_2$ is bijective and continuous, and its inverse f^{-1} is continuous. Homeomorphisms allow for a rich class of deformations as they merely require the topology to stay the same.

As human models and their clothing allow for rather constrained deformations, the most general deformation model considered in this work are *(near-)isometries*. Exact isometries do not allow for surface stretching during deformation, while near-isometries allow for limited stretching. This can be modeled mathematically by requiring the intrinsic geometry of the shape to remain (almost) fixed during the deformation. This deformation model is suitable to model the deformation of both the loco-motion of humans and cloth. Part I introduces this model in more detail and summarizes three of our results: how to robustly compute point-to-point correspondence information using this model, how to detect partial intrinsic symmetries between shapes, and how to leverage this model to spatially align motion sequences of humans in wide clothing.

For human body and body part modeling, higher accuracy can be obtained by employing prior models of shape and deformation behaviour that are specific to the shape class of interest. Unfortunately, due to the high variability of shape both across different subjects and different poses, it is not feasible to express the desired deformations using a manually defined mathematical deformation model. Instead, parametric models for human shape are often learned from a training database of 3D scans that is representative of the desired deformations using a data-driven approach. We employ such parametric models for full human body and human face shape.

Part II builds and applies a parametric model representing human body shape variations due to different identities and body postures. This part summarizes two of our results: how to build such a parametric model from a large database of raw 3D scan data, and how to take advantage of this model to estimate the human body shape and posture for 3D scans captured of people in wide clothing.

Part III studies two parameteric models that represent geometric variations in human face shape across different identities and expressions. This part summarizes three of our results: how to build a global parametric model of high quality using a groupwise approach over the full training data, how to build such a model even in the case where the training data is incomplete or corrupted by noise, and how to build a localized parametric model that allows to represent fine-scale surface detail.

This structure can be seen as focusing on increasingly specific deformations. Part I starts with a near-isometric deformation model that can be used to describe deformations of animal loco-motions and cloth, Part II focuses on the more specific deformations of full human body shape, and Part III finally zooms in on a specific body part, the face. This is illustrated in Figure 1.2.

During the past years, the author has contributed to a number of works that are not discussed in this manuscript due to difference in topic or space constraints (for reference, a list of publications is provided in Appendix A). Especially related to this manuscript are works on modeling deformations of the human tongue [55]. While this work is not discussed in this manuscript, it may allow to include the tongue model in the face models discussed in Part III in the future, which would allow to build a combined model for speech processing. Another related area to the works presented here is the application of our processed human body and face models to two application scenarios. First, we have studied how digital human models can help in the generation of design models for ergonomic design [15]. Second, our developed

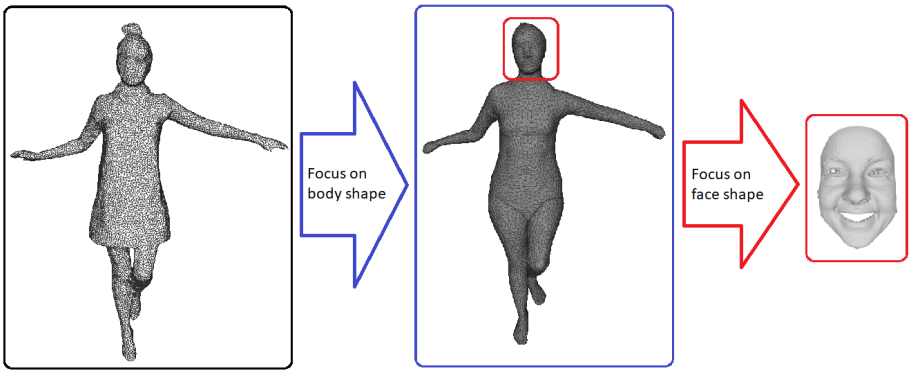

Figure 1.2: *Shapes and deformation models of interest in this work. Part I studies full body shapes with clothing using a near-isometric deformation model, Part II focuses on body shape using a statistical shape space of full bodies, and Part III focuses on face shape using a statistical model of this body part.*

statistical shape spaces have been used to estimate human shapes from partial data [9, 14, 129, 130].

Contributions

The work presented in this manuscript is based on the following main contributions, which have been achieved jointly with colleagues at Saarland University and Max-Planck Institut für Informatik in Germany, and Inria Grenoble Rhône-Alpes in France.

Part I summarizes the following three contributions:

- A method to compute point-to-point correspondence information between two 3D scans of a surface that undergoes near-isometric deformations. This method is particularly robust to acquisition artifacts that changes the perceived topology of the shapes [28].
- A method to compute partial intrinsic symmetry information of a 3D shape. Symmetries are a special case of point-to-point correspondences, mapping from one part of a surface $\mathcal{S}$ to another part of $\mathcal{S}$ [108]. This work was accomplished in the context of the Ph.D. studies of Aurela Shehu.

- A method to compute a dense spatial alignment of motion sequences of humans captured in wide clothing. This method combines the use of a parametric human body model and a near-isometric deformation model [109]. This work was accomplished in the context of the Ph.D. studies of Aurela Shehu and Jinlong Yang.

Part II summarizes the following two contributions:

- A statistical shape space for 3D human body modeling that is built by processing the largest commercially available database of 3D human body scans. This work explores different processing options and presents best practice solutions to build shape spaces [89].
- A method to estimate the shape and posture of a 3D human body model based on scans acquired in wide clothing. This work is especially robust to high levels of occlusion and noise by leveraging motion cues [136]. This work was accomplished in the context of the Ph.D. studies of Jinlong Yang.

Part III summarizes the following three contributions:

- A groupwise approach to build a high-quality parametric shape space for 3D faces that de-couples shape variations caused by identity and expression changes. This method is based on the minimum description length principle [17]. This work was accomplished in the context of the Ph.D. studies of Timo Bolkart.
- A robust version of the previous approach that allows to learn a model from an incomplete training database that may contain shapes that are corrupted by significant noise. This is highly relevant in practice as many existing datasets of 3D faces have a small percentage of data that is missing or contains geometric errors [18]. This work was accomplished in the context of the Ph.D. studies of Timo Bolkart.
- A method to learn a multi-scale shape space for 3D faces that allows to capture localized geometric detail [25]. This work was accomplished in the context of the Ph.D. studies of Timo Bolkart.

Part I

Near-isometric shape modeling

2 Overview

In this part, we study shapes that deform near-isometrically, which means that the surfaces of these shapes allow limited stretching during their deformation. This deformation model is commonly used in computer vision and graphics for two main reasons. First, it allows to model deformations that are relevant in practice such as loco-motions of mammals whose skin does not stretch much during movement, and fairly inextensible cloth often used to manufacture clothing. Second, isometries lead to mathematical models that allow for efficient shape processing, including correspondence computation (e.g. [22, 76, 86, 92]) and shape deformation (e.g. [62, 97]).

Based on the assumption that shapes deform near-isometrically, we have developed algorithms to find features in a way that is invariant to near-isometries [127] and to compute point-to-point correspondences between shapes based on such isometry-invariant feature points [138]. Furthermore, we have considered a particularly interesting case that occurs when working with raw 3D scans of a deforming shape. These scans are typically plagued by geometric and topological acquisition noise due to occlusion and inaccuracies in the 3D reconstruction, which means that the geometry and topology measured by the 3D scanner may locally be different from the geometry and topology of the real object that was scanned. As a consequence, in the acquired data, the near-isometry assumption is often violated. Figure 2.1 illustrates this problem. To address this problem, we have developed an algorithm to compute correspondences for pairs of shapes that is robust to significant noise by considering a set of partial near-isometric mappings instead of a global mapping. This work was published in *Graphical Models* [28] and the motivation for choosing this model and the proposed method are summarized in Chapters 3 and 4, respectively.

This work laid the foundation for two extensions. The first one, published in *ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment 2014* [108] and summarized in Chapter 5, is to compute partial

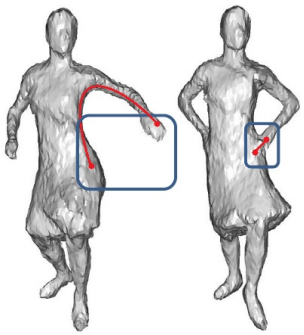

Figure 2.1: *Illustration of acquisition noise. The red curve shows the shortest path along the surface between the two red points. While the red points anatomically correspond on the left and the right model, the shortest path changes drastically due to acquisition noise. Models are reconstructions using space carving [107] of multi-view acquisitions [122].*

intrinsic symmetry information on a single shape $\mathcal{S}$ based on partial near-isometric mappings of different parts of $\mathcal{S}$. The solution of this problem allows for simplified shape manipulation, e.g. by changing the geometry of symmetric parts simultaneously.

The second one, published in *International Conference on 3D Vision 2016* [109] and summarized in Chapter 6, concerns the problem of computing correspondence information of 3D motion sequences of near-isometric shapes. Solving this problem allows to spatially align a 3D motion sequence, i.e. to track the dense geometry of each point on a surface as it deforms over time. The main challenges in solving this problem are robustness and computational complexity; simply applying a correspondence method over pairs of frames either leads to failure cases or is computationally infeasible. Using partial near-isometric mappings, we have developed a method to solve this problem in the special case where the motion sequence represents a human with possibly wide clothing by using information on likely motions of human body shapes as additional constraint.

For more technical details and experimental results on this part, the interested reader is referred to the original publications, which are available in electronic version at

[28]: `https://arxiv.org/pdf/1308.6804.pdf`,

[108]: `http://morpheo.inrialpes.fr/~wuhrer/data/uploads/publications/intrinsic_symmetries_nordia2014.pdf`, and

[109]: `https://hal.inria.fr/hal-01367791/document`.

3 Approximation measures of isometry

When considering near-isometric deformations, the difficulty of finding the mapping between two shapes $\mathcal{S}_1$ and $\mathcal{S}_2$ that leads to the minimum stretching depends strongly on the isometric deformation model that is used. Specifically, the isometric matching considered may be global or local and may concern the entire surface or partial regions. In the following, we review different isometric matching models used in the literature.

Basic definitions

Definition 1. ***Manifolds:*** *Let $\mathcal{M} \subset \mathbb{R}^3$ denote a smooth, orientable 2-manifold embedded in three-dimensional space, possibly with boundaries, which are denoted by $\partial\mathcal{M}$. The orientation of $\mathcal{M}$ is given by outer surface normals $\mathbf{n}(\mathbf{x})$, $\mathbf{x} \in \mathcal{M}$.*

Definition 2. ***Tangent space:*** *Let $T_{\mathbf{x}}\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at point $\mathbf{x} \in \mathcal{M}$. We represent $T_{\mathbf{x}}\mathcal{M}$ using two arbitrary but fixed orthogonal tangent vectors $\mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x})$.*

Definition 3. ***Distances:*** *Let* $\mathrm{dist}_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ *denote the* geodesic distance *between the two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{M}$.*

Definition 4. ***Mappings and isometries:*** *Consider two manifolds $\mathcal{S}_1$ and $\mathcal{S}_2$, and let $f : \mathcal{U} \to \mathcal{S}_2$ denote a mapping from $\mathcal{U} \subseteq \mathcal{S}_1$ to $\mathcal{S}_2$. Let $f_{\mathbf{u}}(\mathbf{x})$ and $f_{\mathbf{v}}(\mathbf{x})$ denote the partial derivatives of f with respect to the tangent space directions $\mathbf{u}$ and $\mathbf{v}$ of $\mathcal{S}_1$ represented in $\mathbb{R}^3$. The first fundamental form $\mathbb{I}_f(\mathbf{x})$ of f at point $\mathbf{x} \in \mathcal{U}$ is*

$$\mathbb{I}_f(\mathbf{x}) = \begin{pmatrix} f_{\mathbf{u}}(\mathbf{x}) \cdot f_{\mathbf{u}}(\mathbf{x}) & f_{\mathbf{u}}(\mathbf{x}) \cdot f_{\mathbf{v}}(\mathbf{x}) \\ f_{\mathbf{v}}(\mathbf{x}) \cdot f_{\mathbf{u}}(\mathbf{x}) & f_{\mathbf{v}}(\mathbf{x}) \cdot f_{\mathbf{v}}(\mathbf{x}) \end{pmatrix}.$$

The function f is an isometry *iff $\mathbb{I}_f(\mathbf{x})$ equals the identity matrix for all $\mathbf{x} \in \mathcal{U}$.*

A function f that is an isometry maintains the geodesic distances between corresponding pairs of points on $\mathcal{S}_1$ and $\mathcal{S}_2$. As the intrinsic geometry of a manifold $\mathcal{S}$ is fully determined by the set of geodesic distances on $\mathcal{S}$, two isometric surfaces have the same intrinsic geometry (but may have distinct extrinsic embeddings in $\mathbb{R}^3$.

Global approximate isometry

Methods that approximate isometry globally consider geodesic distances as global invariants that are consistent up to an error margin $\nu > 0$. That is, two surfaces are considered to be isometric if

$$|\operatorname{dist}_{\mathcal{S}_1}(\mathbf{x}, \mathbf{y}) - \operatorname{dist}_{\mathcal{S}_2}(f(\mathbf{x}), f(\mathbf{y}))| \leq \nu, \forall \mathbf{x}, \mathbf{y} \in \mathcal{U}, \tag{3.1}$$

where $\mathcal{U}$ is typically set to cover $\mathcal{S}_1$. This criterion is sometimes modified to consider relative errors as

$$\max\left(\frac{\operatorname{dist}_{\mathcal{S}_1}(\mathbf{x}, \mathbf{y})}{\operatorname{dist}_{\mathcal{S}_2}(f(\mathbf{x}), f(\mathbf{y}))}, \frac{\operatorname{dist}_{\mathcal{S}_2}(f(\mathbf{x}), f(\mathbf{y}))}{\operatorname{dist}_{\mathcal{S}_1}(\mathbf{x}, \mathbf{y})}\right) \leq (1 + \nu). \tag{3.2}$$

This model strongly constrains possible matching candidates. Isometries are known to be a special case of conformal (i.e. angle-preserving) maps, and hence can be fixed by three pointwise matches on shapes of genus zero [76]. Ovsjanikov et al. [86] have shown that a single point suffices to fix an isometry if the Laplace-Beltrami spectrum of $\mathcal{S}$ is non-degenerate. In this part, we exploit that fixing a point, a tangential direction, and surface orientation is necessary and sufficient to specify an isometric mapping [11].

This model has been used extensively to compute correspondence information between surfaces. A popular approach is to embed the intrinsic geometry of a shape in a canonical space, such that embeddings of isometric shapes become identical [23, 45, 60, 131]. For higher robustness, statistical triangulation algorithms have been applied from a few landmark matches [58, 115, 116] or by voting for approximate solutions [76, 86, 121]. To allow for topological changes, approaches have been proposed that perform a band-limited analysis in the eigenspace of the Laplace-Beltrami operator of $\mathcal{S}$ [7, 24, 77, 92, 112]. In contrast to approaches that embed the intrinsic geometry of $\mathcal{S}$ directly, these approaches successfully handle small topological errors. Sharma et al. [107]

have coupled such an approach with a region growing framework for increased robustness to large topological changes. Another line of work has extended the deformation model to allow for non-isometric distortions, e.g. by using conformal maps with bounds on area distortion [81].

Global approximate isometry in partial regions

The global consistency model is incompatible with partial matching, where pieces of $\mathcal{S}_1$ and $\mathcal{S}_2$ may be missing due to noise and occlusion, because geodesic distances are measured on the complete surfaces. To remedy this, Xu et al. [135] modify the criterion as

$$|\operatorname{dist}_{\mathcal{U}}(\mathbf{x},\mathbf{y}) - \operatorname{dist}_{f(\mathcal{U})}(f(\mathbf{x}),f(\mathbf{y}))| \leq \nu, \forall \mathbf{x},\mathbf{y} \in \mathcal{U}. \tag{3.3}$$

This criterion considers absolute errors and can be modified to consider relative ones as in Equation 3.2. The difficulty that arises from this formulation is that the shortest geodesic paths depend on the shape of the domain $\mathcal{U}$, and hence updating $\mathcal{U}$ influences which pairs of points are mapped approximately isometrically. To remedy this, $\mathcal{U}$ may be restricted to be geodesically convex [135] or $\mathcal{S}_1$ may be restricted to be complete, while $\mathcal{S}_2$ is a deformed possibly incomplete part of $\mathcal{S}_1$ [99]. Partial similarity between shapes can be evaluated using Pareto optimality [21], and this approach aims to find large parts of two surfaces that respect Equation 3.1 in case of near-isometric shape matching. A similar technique can be used to find partial intrinsic symmetries [91].

Local approximate isometry

Another line of work approximates isometry by aiming to maintain the metric tensor in a least-squares-sense. That is

$$\|\mathbb{I}_f(\mathbf{x}) - \mathbf{I}\|_F^2 \leq \nu, \forall \mathbf{x}, \in \mathcal{U}, \tag{3.4}$$

where $\mathbf{I}$ is the identity matrix. This formulation is equivalent to minimizing a matrix norm of the Green deformation tensor. This criterion is local, and can hence readily be used for partial matching. Methods that use this formulation include the preservation of the curvature tensor in the objective function used

for shape matching to encourage maintaining the extrinsic shape [72, 141] or apply the criterion to the Euclidean embedding volume [73, 124].

The problem with this purely local criterion is that the search space of the matching problem becomes too large to be explored by exhaustive search. Hence, it is impractical to find near-isometric mappings between shapes based on only Eqation 3.4. Strategies to reduce the large search space include the use of additional constraints, such as template models [73], temporal coherence [124], or sets of sparse feature correspondences [141]. Another way of local matching is to combine pieces of maps computed using a global near-isometric deformation model obtained by fixing different point-wise matches [64]. When mapping the geometry of a surface to a plane, partial local isometry criteria have been employed [80, 103, 104].

Our model

In summary, using the global approximate isometry criterion on partial regions (Equation 3.3) leads to a complex problem due to the changes in shape of the domain $\mathcal{U}$, and using the local approximate isometry criterion (Equation 3.4) leads to a large search space. To remedy this, we propose a hybrid approach between the two that allows to combine the advantage of global methods of a low-dimensional search space with the advantage of local methods of being able to model partial near-isometric matches. The key idea is to assume that the two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ are related by a global near-isometry in some region $\mathcal{U}$ containing a starting point $\mathbf{s}$. This in turn allows to grow the region $\mathcal{U}$ to the largest size for which the near-isometry assumption holds. This assumption is valid in our scenario, where $\mathcal{S}_1$ and $\mathcal{S}_2$ represent two scans of the same near-isometrically deforming object captured in two poses.

4 Low-dimensional representations for partial near-isometric mappings

This chapter summarizes an efficient algorithm to robustly compute partial correspondence information of near-isometric shapes $\mathcal{S}_1$ and $\mathcal{S}_2$ affected by geometric and topological acquisition noise. The most important contribution is to show that with the matching model discussed in Chapter 3, the search space of this problem is not much larger than in the case of the global isometric problem. Our constructive analysis leads to an efficient algorithm for computing such matches.

Local metric matching

The core of our method is derived from a theorem of differential geometry stating that an isometric map is fixed by one point correspondence and an orthogonal map of the tangent spaces (see e.g. [11, p.201]). For an oriented surface $\mathcal{S}$ representing a (sufficiently) smooth Riemannian manifold, this implies that the correspondence of a point $\mathbf{s} \in \mathcal{S}$ and a local direction $\mathbf{d}_\mathbf{s} \in T_\mathbf{s}\mathcal{S}$ suffices to determine an isometry. For details on smoothness criteria, see [87].

This allows to derive a representation θ as specifying corresponding points $\mathbf{s}_1 \in \mathcal{S}_1, \mathbf{s}_2 \in \mathcal{S}_2$ and tangential directions $\mathbf{d}_{\mathbf{s}_1} \in T_{\mathbf{s}_1}\mathcal{S}_1, \mathbf{d}_{\mathbf{s}_2} \in T_{\mathbf{s}_2}\mathcal{S}_2$ defines a near-isometric mapping between subsets of $\mathcal{S}_1$ and $\mathcal{S}_2$. Using the assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$ are near-isometric, we can start from this information and propagate the correspondence to an infinitesimal neighborhood of $\mathbf{s}_1$ and $\mathbf{s}_2$ by simultaneously walking along corresponding geodesic paths and by matching points reached at the same time. This propagation can then be continued starting from newly matched points in the neighborhood, thereby growing the region $\mathcal{U} \in \mathcal{S}_1$ that is matched. The assumption that $\mathcal{S}_1$ and $\mathcal{S}_2$

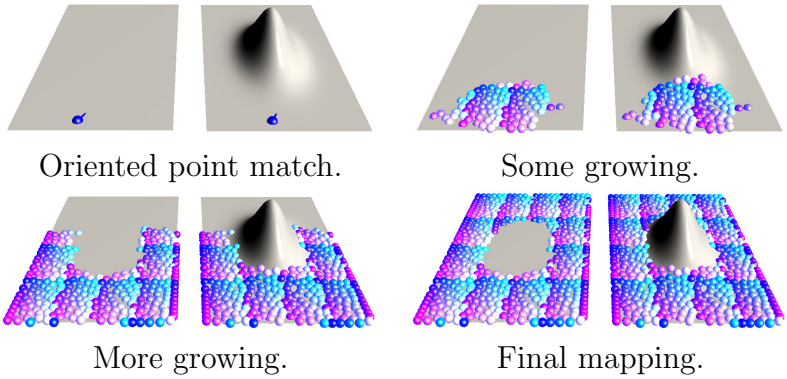

Figure 4.1: *Illustration of the near-isometric region growing process. Corresponding points share the same color. Figure from [28].*

are near-isometric can directly be used to detect the boundary of the largest region $\mathcal{U}$ containing $\mathbf{s}_1$ that can be mapped using a near-isometry by stopping the region growing if a newly added correspondence would lead to a stretching larger than ν. Figure 4.1 illustrated this region growing, where $\mathcal{S}_1$ is part of a plane and $\mathcal{S}_2$ is part of a plane with a hill (this part is not near-isometric to the planar part).

The representation θ is redundant. The first redundancy comes from the choice of $\mathbf{d}_{\mathbf{s}_1}$ in $T_{\mathbf{s}_1}\mathcal{S}_1$, as changing this direction only leads to a rotation in the tangent plane. This redundancy can be removed by starting from an arbitrarily fixed direction $\mathbf{d}_{\mathbf{s}_1}$, by propagating this direction to all points $\mathbf{s}'_1$ of $\mathcal{S}_1$ using parallel transport along an arbitrary but fixed set of geodesics, and by fixing the resulting directions $\mathbf{d}_{\mathbf{s}'_1}$. The second redundancy comes from the choice of $\mathbf{s}_1$ within $\mathcal{U}$. As illustrated in Figure 4.2, for a fixed partial region $\mathcal{U}$, we can replace $\mathbf{s}_1$ in θ by any $\mathbf{s}'_1$ in $\mathcal{U}$ while updating the tangent plane direction and correspondence information without altering the near-isometric mapping.

This implies that the mappings $\mathbf{s}'_1, \mathbf{d}_{\mathbf{s}'_1}, \mathbf{s}'_2, \mathbf{d}_{\mathbf{s}'_2}$ form an equivalence class in θ. After removing one degree of redundancy by fixing the tangential directions on $\mathcal{S}_1$, each mapping f has two remaining degrees of freedom that vary among equivalent maps. Since $\mathcal{U}$ is a manifold, each equivalence class in θ forms a 2-manifold. In practice, we take advantage of this redundancy to make our method robust to noise by starting the propagation from multiple oriented point pairs, by clustering the resulting maps into equivalence classes, and by assembling all maps of a given equivalence class into a single mapping f.

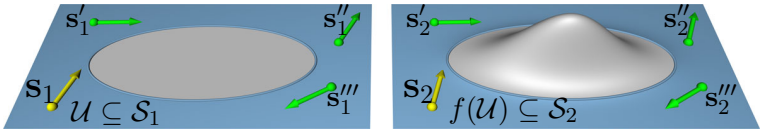

Figure 4.2: *For a given set $\mathcal{U}$ and a corresponding isometry (shown in blue), the choice of the starting point $\mathbf{s}$ is arbitrary. Figure adapted from [28].*

Algorithm

The preceding analysis allows to derive an algorithm that computes and combines consistent partial near-isometric mappings between $\mathcal{S}_1$ and $\mathcal{S}_2$. In our implementation, $\mathcal{S}_1, \mathcal{S}_2$ are represented as point clouds or meshes. The following provides a high-level overview of the algorithm, which is illustrated in Figure 4.3.

The algorithm starts by identifying surface features that can be used to derive oriented point matches. The features are found using an isometry-invariant surface descriptor. Second, features on $\mathcal{S}_1$ and $\mathcal{S}_2$ are matched to identify promising oriented starting point pairs for the region growing. This is achieved using a combination of surface distinctiveness and local descriptor difference. Third, the isometric region growing is started from the set of oriented point pairs that have been identified to be the most reliable. Once the region growing has stopped for the chosen point and direction pairs, the resulting maps are clustered into equivalence classes in θ by discovering manifold structures in the data. For each cluster, the maps are combined into a single coherent mapping f.

The implementation of the presented algorithm is involved as it is not straight forward to account for acquisition noise and to implement geodesic region growing efficiently. Figure 4.4 shows the result of a single partial map found for a pair of shapes, and Figure 4.5 presents a result of a mapping computed for two models where acquisition noise has altered the topology in large contact areas.

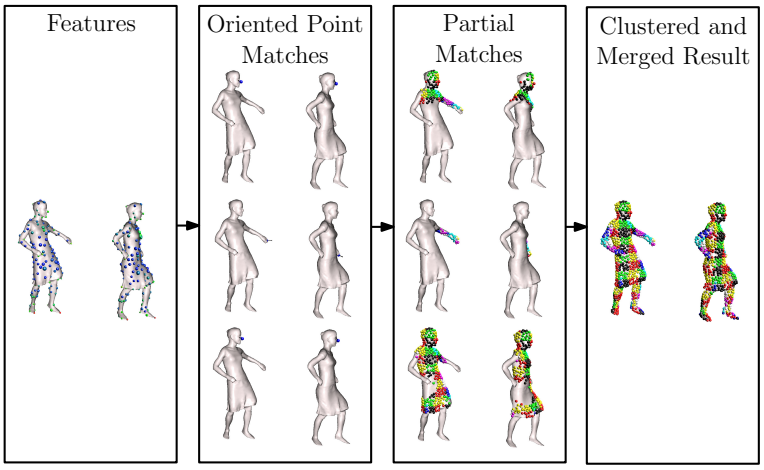

Figure 4.3: *Overview of the pairwise matching pipeline. First, features are detected, second, oriented point matches are computed, third, starting from the oriented point matches, partial isometric matches are found, and finally, different partial matches are clustered and merged. Figure from [28].*

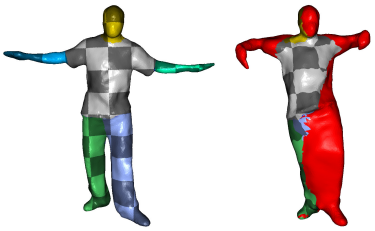

Figure 4.4: *A single partial mapping for an example with large topological changes. The red area is unmatched. Figure from [28].*

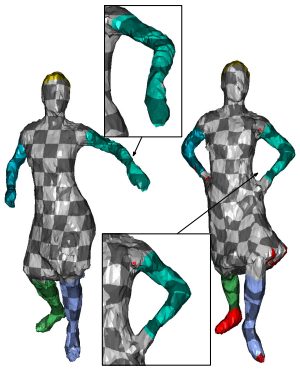

Figure 4.5: *Result of matching $\mathcal{S}_1$ on the left to $\mathcal{S}_2$ on the right. Corresponding points are shown in the same color, and red indicates unmatched areas. Figure adapted from [28].*

5 Detecting partial intrinsic symmetries

The work presented in the previous chapter offers a rich framework that is beneficial for the detection of intrinsic symmetries of a surface. Symmetries play an important role in computer vision and graphics with applications ranging from simplifying the manual design process by modifying shapes in ways that preserve symmetries to shape segmentation and scan completion. In the literature, both extrinsic and intrinsic symmetries have been studied, and more recently, some works have focused on symmetries of parts of surfaces. We refer to Mitra et al. [82] for an extensive survey of this topic.

This chapter focuses on intrinsic symmetries, and in this case, two points $\mathbf{s}_1, \mathbf{s}_2$ of shape $\mathcal{S}$ are (globally) symmetric if there is a mapping f with $f(\mathbf{s}_1) = \mathbf{s}_2$ such that Equation 3.1 holds with $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$, $\mathcal{U} = \mathcal{S}$ and $\nu = 0$. We consider partial symmetries. However, if we allow arbitrarily small parts $\mathcal{U}$, for a sufficiently smooth surface, every point of $\mathcal{S}$ becomes intrinsically symmetric to every other point of $\mathcal{S}$. Furthermore, as in Chapter 4, we are interested in relaxing the notion of isometric mappings to near-isometric mappings. Hence, we consider (ν, r)-symmetries in the following:

Definition 5. *Two points $\mathbf{s}_1$ and $\mathbf{s}_2$ of $\mathcal{S}$ are called (ν, r)-symmetric if there is a mapping function $f : \mathcal{U} \to \mathcal{S}$ with $f(\mathbf{s}_1) = \mathbf{s}_2$, where $\mathcal{U}$ is a geodesic disk of radius r centered at $\mathbf{s}_1$, such that Equation 3.1 is satisfied with $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$.*

It is interesting to note that while (ν, r)-symmetries form an equivalence relation for $\nu = 0$, this no longer holds for $\nu > 0$. However, since we only choose $\nu > 0$ to counteract acquisition noise, we assume that there is an underlying equivalence class for the symmetries, and report this by post-processing the detected symmetries.

Definition 5 shows that the framework developed in Chapter 4 can be used

to detect symmetries by finding near-isometric self-mappings of $\mathcal{S}$ from one region to another. However, a major difficulty arises for intrinsic symmetries, as it is believed that the space of partial intrinsic symmetries of a shape might be too large to explore with the goal to report all symmetries [91, 135]. This implies that directly mapping feature points will either result in a slow algorithm or miss some of the (ν, r)-symmetries.

The reason for the large search space is the presence of infinitesimal symmetries. An *infinitesimal symmetry* occurs is f encodes infinitesimal movements of the points $\mathbf{x} \in \mathcal{U}$ that lead to (ν, r)-symmetries. Infinitesimal symmetries are also called continuous symmetries in the literature since the mappings involved form a continuous set. In contrast, when f encodes movements by fixed geodesic distances of the points $\mathbf{x} \in \mathcal{U}$ that lead to (ν, r)-symmetries, we call the resulting mapping a *discrete symmetry.*

Our method proceeds by first detecting and factoring out infinitesimal symmetries. Points $\mathbf{s} \in \mathcal{S}$ that have infinitesimal symmetries have the property that the geodesic disk $\mathcal{U}$ of radius r centered at $\mathbf{s}$ can move along at least one (high-dimensional) direction $\mathbf{d}$ along surface $\mathcal{S}$ while remaining (ν, r)-symmetric. In this case, we call $\mathbf{d}$ a slippage direction. It is known that there can be at most three linearly independent slippage directions [95], and we can therefore classify points on $\mathcal{S}$ as k-slippable for $k = 0, 1, 2, 3$. We have developed an algorithm to classify points on $\mathcal{S}$ according to their slippability by constructing a matrix that has the slippage directions as eigenvectors with corresponding eigenvalue zero. This is an extension of prior works on extrinsic symmetries of shapes [47] that is not straightforward due to the increase of the number degrees of freedom of the movement of $\mathcal{U}$ in the intrinsic case.

In the subsequent analysis, slippable points are excluded from consideration. Focusing on the set of 0-slippable points only allows to exhaustively search and report the space of discrete symmetries using the region-growing algorithm presented in Chapter 4. Figure 5.1 provides an illustration of this process. Given as input the cube shape in this case along with a radius of interest r and a tolerance ν, we first determine point slippability. Figure 5.1 shows the $0, 1, 2$ and 3-slippable points in white, red, green and blue, respectively. Note that all 0-slippable points are concentrated around the corners of the cube. This allows to exhaustively search and report all discrete symmetries between 0-slippable points.

For increased efficiency, we only match point and direction pairs that have similar intrinsic spectral descriptors (we choose the wave kernel signature for this purpose [7]). Once all discrete (ν, r)-symmetries have been found,

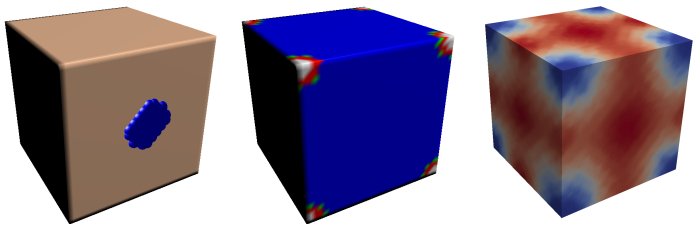

Figure 5.1: *Illustrative example. Left to right: size of r, point slippability, symmetry-factored embedding distance from a corner of the cube. Figure from [108].*

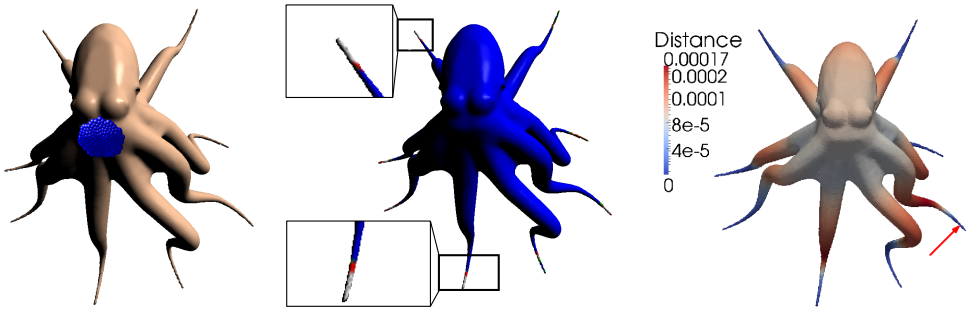

Figure 5.2: *(a) Symmetry size r shown on the shape of interest. (b) Point slippability for threshold 0.03 times the intrinsic area of the shape. The $0, 1, 2$ and 3-slippable points are shown in white, red, green and blue, respectively. (c) Visualization of the equivalence class of the marked point. The symmetry-factored distance [75] is shown, i.e. points at distance 0, shown in blue, are (ν, r)-symmetric to the marked point, while points at larger distances have larger intrinsic distortions in an r-neighborhood. Figure adapted from [108].*

we cluster them to find equivalence classes by extending symmetry-factored embeddings and distances [75] to our scenario. This allows to visualize the detected discrete symmetries in an intuitive way.

Figure 5.2 visualizes a result of both the infinitesimal and discrete (ν, r)-symmetries. Note that for the octopus shape, the tips of the tentacles are detected as non-slippable and belong to the same equivalence class of (ν, r)-symmetries.

6 Spatially aligning 3D sequences of clothed humans

This chapter considers the problem of computing correspondence information of 3D motion sequences of near-isometric shapes. Solving this problem in general is very challenging due to the rich deformations permitted by shapes that undergo fast deformations (compared to the frame-rate of the acquisition system) and by significant noise present in acquisitions of 3D motion sequences. Greedily applying pairwise correspondence methods as the one presented in Chapter 4 between consecutive frames is known to lead to problems due to drift.

State-of-the-art methods that rely strongly on a near-isometry assumption can only handle limited topological noise [114]. Alternative methods rely strongly on temporal coherence and have problems following fast motions [29]. To remedy this, most works that study the spatial alignment of human motion sequences employ template shapes as prior [1, 2, 39, 111]. While this allows for robust algorithms, the shape of the body and clothing are strongly constrained. Recent general methods require less temporal consistency and allow for topology changes by changing the topology and geometry of the shape over time [43], which leads to impressive results, but no longer respects our model of a fixed underlying topology that deforms in a near-isometric manner.

We propose to compute the point-to-point correspondence information across the sequence by growing near-isometric patches starting with a set of reliably obtained body landmarks. The resulting correspondence is more robust than methods that rely on local features or temporal coherence for matching.

Figure 6.1 gives an overview of our pipeline. To start, we automatically select one frame of the sequence as a template $\mathcal{T}$. Recall our assumption

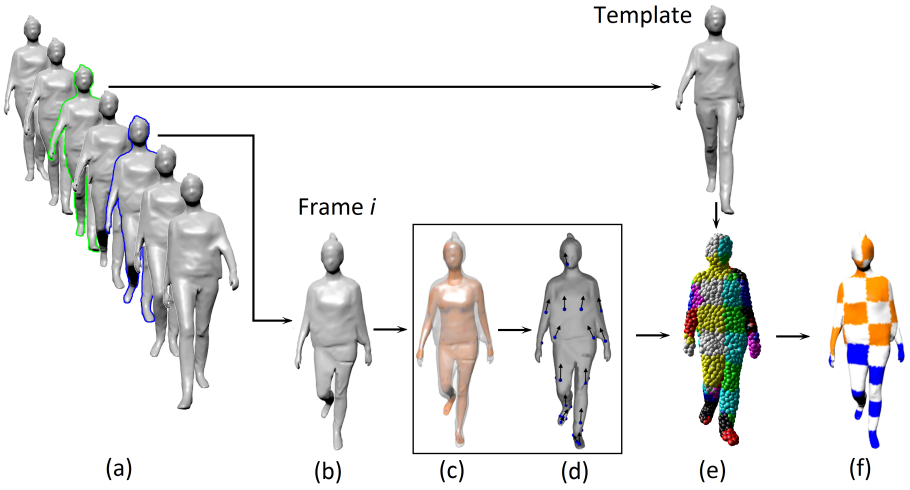

Figure 6.1: *Method overview. Given an input sequence of 3D meshes shown in (a), each frame is processed in three steps, which are shown for the frame indicated in blue and shown in (b). First, a statistical model of undressed body shape is fitted to the frame (c). Pre-marked anatomical points on the fitted human body model are mapped to the input frame (d). Second, these anatomical markers are used to guide a partial near-isometric correspondence computation between an automatically selected template (green model in (a), and shown in top row) and the frame (e). Third, the resulting correspondences are used as assignments to deform the template to the input frame (f). Figure from [109].*

that the scene has a fixed topology and deforms near-isometrically, and that observed topology changes are caused by acquisition artifacts. We further assume that the acquired sequence contains at least one frame where the observed topology corresponds to the true one. As our goal is to select as $\mathcal{T}$ a frame with the correct topology, we choose a frame with maximal number of components and minimal genus. Of all these frames, we choose the one that has the minimum area ratio of its point areas in order to have a good sampling of the geometry of $\mathcal{T}$.

After selecting $\mathcal{T}$, each frame $\mathcal{S}$ is processed in three steps. First, we fit a statistical model of an undressed body to $\mathcal{S}$, as explained in Chapter 10 (see Figure 6.1(c)). This body estimate serves to project a set of oriented anatomical markers pre-defined on the statistical model to the surface of $\mathcal{S}$, which allows to define oriented point pairs between all pairs of frames, and in particular, between $\mathcal{T}$ and $\mathcal{S}$ (see Figure 6.1(d)). Second, the resulting oriented point pairs are used to compute a set of point-to-point correspondences between $\mathcal{T}$ and $\mathcal{S}$ using the partial near-isometric region growing framework

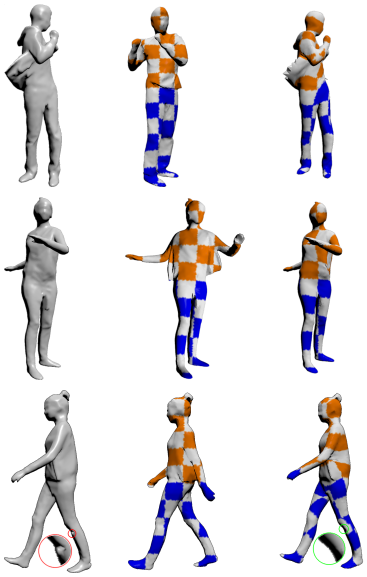

Figure 6.2: *Alignment result on three sequences. From left to right: $\mathcal{S}, \mathcal{T}$ and our result. Figure adapted from [109].*

introduced in Chapter 4 (see Figure 6.1(e)). While this gives a spatial correspondence from $\mathcal{T}$ to every frame of the sequence, not every point of a particular frame $\mathcal{S}$ may have a correspondence due to acquisition noise. To obtain a consistent topology that deforms over time, we therefore post-process the resulting correspondences in a third step by deforming $\mathcal{T}$ to $\mathcal{S}$ using a template fitting technique [3] with the previously computed correspondences as assignments.

This pipeline allows to compute correspondence information for sequences of human bodies captured in wide clothing, as shown in Figure 6.2.

Part II

Human body modeling

7 Overview

In this part, we focus on the analysis of 3D human body scans. Human body models are especially important in various applications, e.g. the immersion in virtual worlds or the design of ergonomic products. As human body shape varies significantly across subjects and body postures, it is not feasible to directly define a suitable and tractable mathematical deformation model as was done in Part I. Instead, generative models of human body shape are commonly learned from databases of example scans exhibiting a variety of different subjects, postures, and recently even motion patterns [3, 6, 54, 78, 90]. This leads to shape spaces that typically aim to compactly represent body shape and to de-couple shape changes due to different identity, different body posture, and if modeled dynamic effects. Chapter 8, which is based on our previous review [26], provides a proper definition of shape spaces and gives an overview of existing generative models for dense 3D human body modeling.

To build such generative models, the raw scans present in the training set are first registered by computing point-to-point correspondences between the scans. Our results include the fully automatic computation of point-to-point correspondences between large databases of static human body models with the help of template shapes that capture the typical anatomy of a human body and are deformed using a skeleton-based deformation. When fitting such a template shape to raw noisy 3D scan data, the main challenge is the initialization of the position and posture of the template with respect to the data. We have developed one of the first robust methods to allow for automatic initialization by predicting and subsequently using the locations of anatomic landmarks on the raw scan data, which allowed processing over 4000 human body scans automatically by utilizing tools from machine learning [132]. More recently, we have extended this work to build a family of high-quality generative body models that are both compact and general. This work was published in *Pattern Recognition* [89], and Chapter 9 summarizes this work.

A generative body model allows to significantly reduce the search space

for subsequent correspondence problems to the point where large amounts of noise and occlusion can be handled. A particular research focus of mine has been to estimate the human body shape in motion for different morphologies, different motions, and different clothing styles. We recently used our developed generative body models to improve our previous result [128] to the point where body shape is reliably estimated under wide and layered clothing. This is challenging due to acquisition noise and high levels of occlusion caused by the wide clothing. Our automatic method to solve this problem takes advantage of visual cues observed during motion. In particular, we take advantage of the fact that the identity of the sought body shape is fixed over the sequence, and that the body shape in motion is contained inside the observed 3D model (showing the dressed person) at every captured time instant. This work was published in *International Conference on 3D Vision* [136], and Chapter 10 summarizes this work.

For more technical details and experimental results on this part, the interested reader is referred to the original publications, which are available in electronic version at

[89]: `https://hal.inria.fr/hal-01136221v2/document` and

[136]: `https://hal.inria.fr/hal-01344795v4/document`.

8 Statistical priors for 3D human body modeling

There has been an extensive amount of work in computer vision and graphics on modeling humans and their actions and interactions. Here, we focus our attention on modeling *dense 3D human body shape*, possibly in clothing. We further restrict ourselves to *statistical deformation models* derived in a data-driven way. The main reason to use such statistical shape models instead of template shapes that are deformed using a regularization constraint, is that statistical deformation models are typically low-dimensional, thus allowing to reduce the search space when these models are used in optimization tasks, such as the reconstruction of a shape from partial, noisy, or occluded data.

Basic definitions

As deformation model of human body shapes, we aim to find a statistical shape space that informatively represents the shape variations of human bodies using a low-dimensional parameter space. In this work, we learn such a shape space from a dataset of 3D scans showing the shape variations of interest. To leverage a dataset of 3D scans for training, a notion of distance between two scans is required and it is not straight forward to measure the distance between two raw scans that consist of a different number of unstructured vertices. The most common way to proceed is to first pre-process the database by registering all scans to a common model, thereby computing anatomic point-to-point correspondence information between the scans. In the following, let $\mathcal{S}$ denote the surface of such a registered 3D model consisting of n vertices, and let $\mathbf{c} \in \mathbb{R}^{3n}$ denote its associated coordinate vector obtained by stacking the vertex coordinates of $\mathcal{S}$ in an arbitrary but fixed order. Computing distances between two shapes $\mathcal{S}_1$ and $\mathcal{S}_2$ can now be

achieved by computing the difference between $\mathbf{c}_1$ and $\mathbf{c}_2$ after rigidly aligning $\mathcal{S}_1$ and $\mathcal{S}_2$ in $\mathbb{R}^3$.

We follow Dryden and Mardia [44] and define the *shape* of $\mathcal{S}$ as its geometric information after removal of effects caused by translation, rotation, and optionally uniform scaling. In our context, a shape space is traditionally defined as the set of all possible shapes consisting of n vertices in $\mathbb{R}^3$. However, as we are only interested in a particular class of shapes (namely ones representing human models), we use the term *shape space* to refer to a d-dimensional parameter space with $d << 3n$ that allows to represent objects of this class. That is, each object is represented by a vector $\mathbf{w} \in \mathbb{R}^d$.

Definition 6. *We define a* statistical shape space *of a class of 3D models as a shape space with a probability distribution, called* prior *and expressed using a density function $f(\mathbf{w})$, that can be used to measure how likely it is that an object of the given class would have a particular parametric representation* $\mathbf{w}$ *in shape space.*

In this way, we use statistical shape analysis as *generative* technique. The surface $\mathcal{S}$ is represented by d coefficients, which form a vector $\mathbf{w} \in \mathbb{R}^d$. A generator function

$$\mathbf{F}(\mathbf{w}) : \mathbb{R}^d \rightarrow \mathbb{R}^{3n} \tag{8.1}$$

allows to generate the coordinate vector $\mathbf{c}$ from these coefficients.

In our work, we formulate the variations of human body and face shapes as bilinear models that separate variation due to identity and pose (i.e. posture in the case of bodies and expression in the case of faces). In this formulation, the representation $\mathbf{w}$ can be decomposed into two parameter vectors $\mathbf{w}_{id} \in \mathbb{R}^{d_{id}}$ and $\mathbf{w}_{pose} \in \mathbb{R}^{d_{pose}}$ with $d = d_{id} + d_{pose}$, and the function $\mathbf{F}(\mathbf{w})$ takes the following form

$$\begin{aligned} \mathbf{c}\left[l\right](\mathbf{w}_{id}, \mathbf{w}_{pose}) &= \left(\sum_{i=1}^{d_{id}} \sum_{j=1}^{d_{pose}} a_{ijl} \mathbf{w}_{id}[i] \mathbf{w}_{pose}[j] \right) \\ &+ \left(\sum_{i=1}^{d_{id}} a_{il}^{(id)} \mathbf{w}_{id}[i] \right) \\ &+ \left(\sum_{j=1}^{d_{pose}} a_{jl}^{(pose)} \mathbf{w}_{pose}[j] \right) \\ &+ o_l, \end{aligned} \tag{8.2}$$

where $\mathbf{c}[l], \mathbf{w}_{id}[i]$ and $\mathbf{w}_{pose}[j]$ denote the l-th, i-th and j-th components of $\mathbf{c}, \mathbf{w}_{id}$ and $\mathbf{w}_{pose}$, respectively, and where the scalar weights $a_{ijl}, a_{il}^{(id)}, a_{jl}^{(pose)}$ and offsets o_l are learned from training data.

The priors used for these types of models vary depending on the application and precise bilinear models used. Common priors include multidimensional box priors that clamp the parameter space to allowed values, which occur with uniform probability. This model is most commonly used for $\mathbf{w}_{pose}$ in case of human body shape, as the parameter space often corresponds to joint angle representations. Other commonly used priors include multivariate normal distributions for $\mathbf{w}_{id}$ for human bodies and faces, and multivariate normal distributions for $\mathbf{w}_{pose}$ for faces.

Shape spaces for human bodies

For 3D human body shapes in a standardized posture, Allen et al. [3] proposed a statistical shape space by performing principal component analysis (PCA) over a registered set of training data acquired in similar posture. This model has been used to predict 3D human body shape from partial information such as one or more images or measurements (e.g. [14, 33, 106]). It has also been extended to part-based models by performing PCA on different segments of the body independently [134, 144].

Multiple statistical shape spaces have been introduced to model shape variation due to both different identities and postures. Anguelov et al. [6] introduced the *SCAPE model*, which learns a PCA shape space for body shape variations due to different identities from subjects in different postures. SCAPE further learns a mapping from skeleton-based posture parameters to shape changes using a database containing one subject in multiple poses. The model then uses the assumption that the shape changes controlled by the two factors are decorrelated to combine the two effects. While this is a simplistic assumption, SCAPE models human body shape successfully and has been used for a variety of applications (see e.g. [8, 52, 126, 142]). Training the SCAPE model has been improved by learning the model while simultaneously optimizing the correspondence information of the training data [57]. SCAPE has further been extended to model shape changes caused by dynamic effects of motion [90, 120] and to model layers of clothing [51].

Some works allow for posture variation by modeling shape changes caused

by identity and posture as correlated. This can be achieved by performing PCA on a rotation-invariant encoding of the shapes [54] or by using multilinear models either globally [53] or on body parts [34].

While modeling the correlation between shape changes caused by identity and posture is a good model, it leads to representations that are difficult to optimize directly. Hence, methods based on SCAPE remain popular. SCAPE-like models represent the 3D shape as triangle mesh, and use a set of transformations per triangle to encode shape variations. This requires solving a computationally demanding optimization problem to convert between a shape space representation and vertex coordinates of a shape. To overcome this, recent works have proposed multilinear shape spaces that operate on vertex coordinates directly [59, 78, 84]. While this allows for faster reconstructions of coordinate vectors, these shape spaces generally encode less of the shape variation present in the training database.

Recently, the field of deep learning has started to influence 3D human body modeling [42, 48]. However, to the best of my knowledge, there are currently no representations of the dense 3D geometry of human body models that are directly learned via deep learning.

Context

In our works, we have experimented with different statistical shape spaces for human body modeling. To allow for efficient optimization, we used a linear PCA model on the coordinate representation $\mathbf{c}$ for shape changes caused by varying identities coupled with a skeleton-based linear blend skinning deformation for posture changes. This model has also been used in the literature [59] and is called simplified SCAPE (S-SCAPE) in the following. The S-SCAPE representation allows to efficiently fit the shape space to an input scan. We have used this model for the fully automatic registration of raw scans in different postures [132]. The main problem we observed with this shape space are deformation artifacts that occur for two reasons. First, linear blend skinning is a simple model that is known to cause localized artifacts in case of strong joint rotations. Second, PCA models that encode identity variation are typically trained on a set of registered 3D scans of people acquired while asked to maintain a "standard posture", which implies that they still encode small posture variations caused by anatomical differences (such as subjects naturally leaning forward or backward while standing). These

posture variations subsequently interfere with the skeleton-based deformation and may cause errors affecting large areas of the body.

To address these problems, we have experimented with posture-invariant shape spaces [133] that remove the effect of posture changes. This encoding allows to change the posture of one given body shape $\mathcal{S}_1$ to that of another $\mathcal{S}_2$. Furthermore, this shape space allows to reliably estimate body shape under clothing based on one or multiple scans of a subject [128]. While this shape space reduces the artifacts discussed above, its main disadvantage is that generating a 3D shape from the parameter vector $\mathbf{d}$ in shape space requires solving an optimization problem, leading to high computation times.

To combine the advantages of both approaches, in the following we consider shape spaces based on S-SCAPE that learn the PCA space encoding variations due to identity on posture-normalized versions of the registered scans. This allows for efficient optimizations in shape space while reducing artifacts caused by posture variations in the training data. Two other research groups have recently proposed models that are similar in spirit [78, 84].

9 Statistical shape spaces for 3D human modeling

This section summarizes our work on rebuilding a statistical S-SCAPE shape space of 3D human bodies that is commonly used in the community because of its effectiveness to model the shapes of interest and its efficiency. Our contribution are not entirely new methods, but rather the application of best practices and solutions for the automatic processing of 3D body scans to the largest commercially available scan database. This preprocessing of thousands of 3D scans for learning the model is challenging, and we use robust techniques to this end.

S-SCAPE model

We start with a more formal description of the S-SCAPE model [59] by providing the specific generator function of the form presented in Equation 8.2 that is used by S-SCAPE. To simplify the notation, we consider shapes represented in homogeneous coordinates in this chapter, i.e. $\mathbf{c}$ is represented in $\mathbb{R}^{4n}$.

S-SCAPE is learned using preprocessed training shapes $\mathcal{S}_i$ captured in a standard posture $\mathbf{w}_{pose,0}$ using PCA. This allows to represent each coordinate vector used for training with a parameter vector $\mathbf{w}_{id}$ and to generate new body shapes in posture $\mathbf{w}_{pose,0}$ as

$$\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose,0}) = \mathbf{A}\mathbf{w}_{id} + \bar{\mathbf{c}}, \tag{9.1}$$

where $\mathbf{A} \in \mathbb{R}^{4n \times d_{id}}$ is the projection matrix computed by PCA and $\bar{\mathbf{c}} \in \mathbb{R}^{4n}$ is the mean body shape of the training data.

To enable posture variation, an articulated skeleton is fitted to the mean

human body shape and a linear blend skinning model attaching the surface to the bones is defined. This allows to generate a fixed identity $\mathbf{w}_{id,0}$ in arbitrary posture as

$$\mathbf{c}_i(\mathbf{w}_{id,0}, \mathbf{w}_{pose}) = \sum_{j=1}^{b} \omega_{ij} \mathbf{T}_j(\mathbf{w}_{pose}) \mathbf{c}_i(\mathbf{w}_{id,0}, \mathbf{w}_{pose,0}), \tag{9.2}$$

where $\mathbf{c}_i$ denotes the homogeneous coordinates corresponding to the i-th vertex of $\mathbf{c}$, b is the number of bones in the skeleton, $\mathbf{T}_j \in \mathbb{R}^{4\times4}$ is the transformation of the j-th bone and ω_{ij} are rigging weights.

Combining Equations 9.1 and 9.2 allows to reconstruct an arbitrary identity $\mathbf{w}_{id}$ in an arbitrary posture $\mathbf{w}_{pose}$ as

$$\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose}) = \mathbf{T}(\mathbf{w}_{pose})\mathbf{A}\mathbf{w}_{id} + \mathbf{T}(\mathbf{w}_{pose})\bar{\mathbf{c}}, \tag{9.3}$$

where $\mathbf{T}(\mathbf{w}_{pose}) \in \mathbb{R}^{4n\times4n}$ is a block-diagonal matrix with the per-vertex transformations on the diagonal. Note that this model allows for very efficient reconstructions at the cost of not representing fine-scale detail such as muscle-bulging.

Data processing

We build a S-SCAPE model by combining the largest commercially available database of human body scans [96] with a database that is publicly available [54] for a total of about 4500 scans of North American and European adults, thereby modeling a large variation of adult body shape. The data processing consists of the computation of accurate point-to-point correspondences between the raw scans and the normalization of their body postures. To simplify notation, in the following discussion, we use $\mathcal{S}_i$ to represent the shape of both the raw and the processed scan as the meaning is clear from the context.

Non-rigid template fitting To compute point-to-point correspondence information between the scans, we follow a template fitting method [3] that aligns the geometry of a template shape $\mathcal{T}$ representing a typical body shape to each raw scan $\mathcal{S}_i$. The template fitting transforms each vertex using an affine transformation by optimizing three energy terms. First, a data term is used that aims to deform the geometry of $\mathcal{T}$ to be close to $\mathcal{S}_i$. Second,

to regularize the overall body shape, the deformation is required to stay locally smooth by transforming neighboring points on $\mathcal{T}$ using similar affine transformations. Third, reliable landmarks present in the databases are used to guide the deformation by encouraging corresponding anatomical markers to be close-by on $\mathcal{T}$ and $\mathcal{S}_i$.

Initialization The result of the non-rigid template fitting depends strongly on the initial alignment of $\mathcal{T}$ and $\mathcal{S}_i$. We have explored two initialization strategies. The first one uses a simple static template shape, while the second one uses an existing S-SCAPE model learned from much fewer scans [59] to adjust the shape and posture of $\mathcal{T}$ to each $\mathcal{S}_i$. The second initialization results in better fitting accuracy.

Bootstrapping After the non-rigid template fitting, the fitted mesh may still be far from the raw scan in cases where the body shape of $\mathcal{S}_i$ is far from $\mathcal{T}$. A remedy is to visually examine each fitting, discard fittings of low quality, and learn a S-SCAPE space using the samples that passed the visual inspection. This S-SCAPE space is then used as initialization to perform a fitting during the next pass. We found that performing the bootstrapping for a few iterations results in almost all registered scans passing the visual inspection.

Posture normalization As a final preprocessing step, we normalize the posture of all registered scans $\mathcal{S}_i$ to avoid modeling posture variation in PCA space. This is required even though the models are capture in a standard posture, as small variations in arm posture are present. We explored two methods for posture normalization. The first uses a skeleton model and Laplace surface deformation to process the scans [84], while the second method based on our previous work uses localized Laplace coordinates to this end [133]. We found that both ways of performing posture normalization significantly improve the resulting shape space.

Figure 9.1 visualizes the resulting S-SCAPE spaces. The first three components already capture rich shape variations, such as variations in height, body weight, and from male and female body shape. Note that variation due to body posture such as small arm motion are present in the first row, but removed in the two bottom rows thanks to posture normalization.

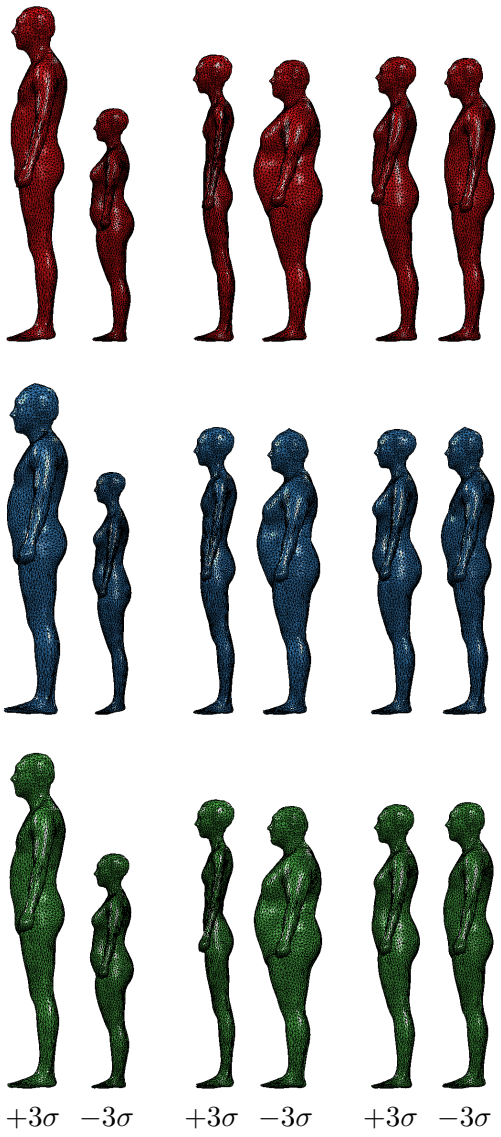

Figure 9.1: *Visualization of the first three PCA eigenvectors scaled by $\pm 3\sigma$ (standard deviation). Shown are the S-SCAPE spaces trained using our preprocessed data without (row 1) and with posture normalization using [130] (row 2) and [84] (row 3). Figure adapted from [89].*

10 Estimating body shape and motion under wide clothing

The S-SCAPE space learned in the previous chapter can directly be applied to estimate the body shape in motion based on an acquired motion sequence showing a person in wide clothing. This is important in a variety of applications ranging from virtual change rooms, creating customizable avatars for games, and security. Automatically estimating body shape under loose clothing is challenging due to high surface variations and significant body occlusion caused by the clothing.

We propose a solution to this problem that fits a S-SCAPE model to the observed sequence of raw 3D scans. That is, for a sequence of m frames $\mathcal{S}_1, \ldots, \mathcal{S}_m$ showing a dressed person in motion, we solve for one identity parameter $\mathbf{w}_{id}$ and m posture parameters $\mathbf{w}_{pose,1}, \ldots, \mathbf{w}_{pose,m}$ such that $\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose,i})$ defined by Equation 9.3 is close to $\mathcal{S}_i$. More specifically, we solve for $\mathbf{w}_{id}, \mathbf{w}_{pose,i}$ by optimizing an energy of the form

$$\begin{aligned} E_{sequence} = & \sum_{i=1}^{m} (\omega_{lnd} E_{lnd}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i) \\ + & \quad \omega_{data} E_{data}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i) \\ + & \quad \omega_{cloth} E_{cloth}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)), \end{aligned} \tag{10.1}$$

where $E_{lnd}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$, $E_{data}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$ and $E_{cloth}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$ are energy terms weighted by scalars $\omega_{lnd}, \omega_{data}$ and ω_{cloth}. The landmark term $E_{lnd}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$ measures the distance between a set of automatically computed landmarks on $\mathcal{S}_i$ and their corresponding anatomical points on $\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose,i})$. This energy is crucial to obtain a rough estimate of the body shape and posture at each frame. The data term $E_{data}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$ measures the distance of points $\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose,i})$ to the nearest neighbors on scan $\mathcal{S}_i$ and serves to pull the estimate towards the observed scan surface. This term is crucial to obtain a good estimate for $\mathbf{w}_{id}$. The clothing

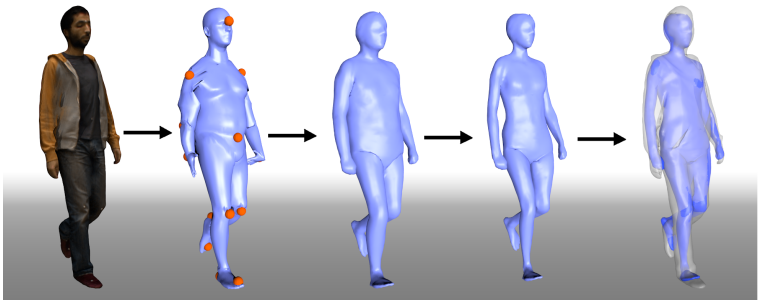

Figure 10.1: *Overview of the proposed pipeline. From left to right: input frame, automatically computed landmarks using [144], result after estimation of initial identity and posture, final result, and overlay of input and final result. Figure from [136].*

term $E_{cloth}(\mathbf{w}_{id}, \mathbf{w}_{pose,i}, \mathcal{S}_i)$ accounts for the loose clothing by encouraging $\mathbf{c}(\mathbf{w}_{id}, \mathbf{w}_{pose,i})$ to be located inside the observation $\mathcal{S}_i$. Since the cloth term is applied to all frames, it allows to take advantage of the motion cues observed throughout the sequence as the clothing moves close to localized regions of the body in different frames.

Optimizing energy 10.1 with respect to $\mathbf{w}_{id}, \mathbf{w}_{pose,1}, \ldots, \mathbf{w}_{pose,m}$ jointly is not feasible for motion sequences containing hundreds of frames. To solve this problem, we have derived an optimization schedule that solves three steps in order. First, $\mathbf{w}_{id}$ is initialized by solving Equation 10.1 for a small subsequence containing the first few frames of the sequence. We call the result of this optimization $\mathbf{w}_{id,0}$. Second, posture parameters are optimized sequentially by minimizing Equation 10.1 for one frame at a time with respect to $\mathbf{w}_{pose,i}$ while fixing $\mathbf{w}_{id,0}$. Third, the identity is refined by optimizing Equation 10.1 with respect to $\mathbf{w}_{id}$ wile fixing $\mathbf{w}_{pose,i}$. During the three steps of the optimization, the $\omega_{lnd}, \omega_{data}$ and ω_{cloth} in Equation 10.1 follow a carefully chosen weight schedule. Figure 10.1 provides an overview of the method.

To evaluate the proposed framework, we have captured a database of six subjects (three male, three female) performing three different motions and wearing three clothing styles each using the Kinovis mulit-camera platform at Inria Grenoble Rhône-Alpes. This dataset has already been used by the computer vision community in follow-up works (e.g. [140]). Figure 10.2 shows six representative frames of the database. Figure 10.3 evaluates the posture estimation on manually placed markers for walking sequences captured in tight clothing. Note that the use of S-SCAPE as statistical prior prevents drift. Figure 10.4 shows further qualitative results of our method.

Figure 10.2: *Six representative examples of frames of our motion database. From left to right, a female and male subject is shown for tight, layered, and wide clothing each. Figure from [136].*

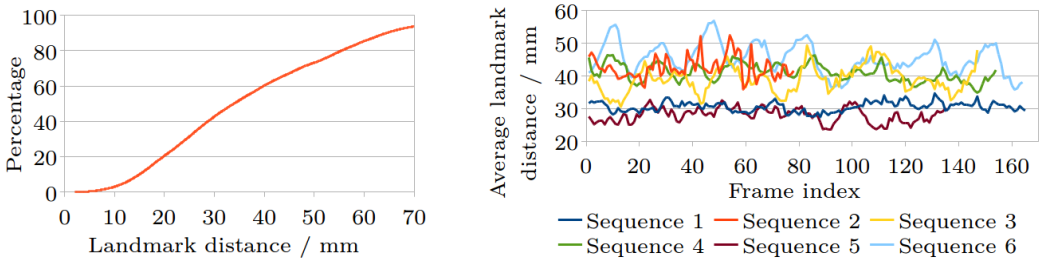

Figure 10.3: *Accuracy of posture estimation over the walking sequences of all subjects in tight clothing. Left: cumulative landmark errors. Right: average landmark error throughout each sequence. Figure from [136].*

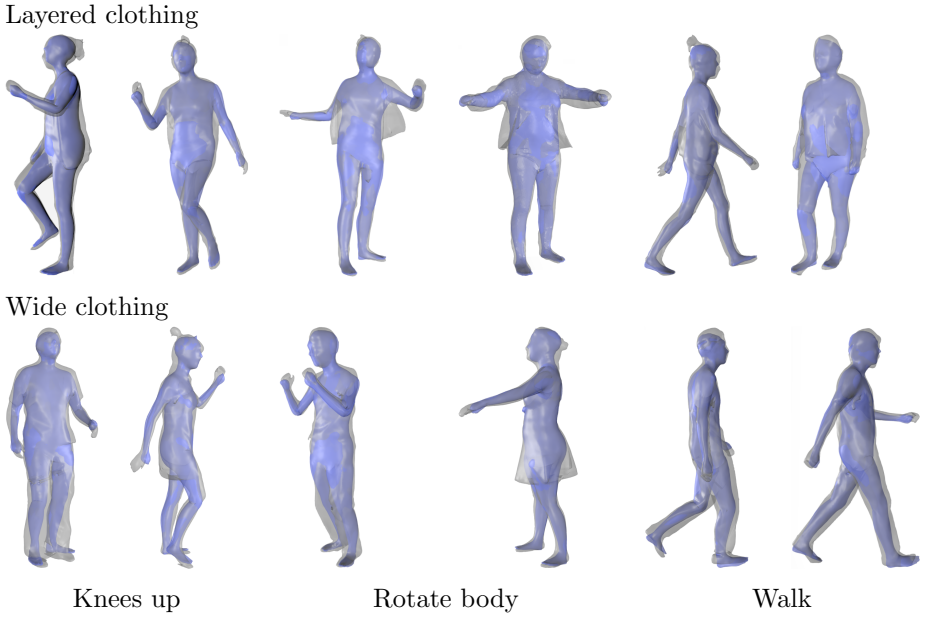

Figure 10.4: *Overlay of input data (grey) and our result (blue). Figure adapted from [136].*

Part III

Human face modeling

11 Overview

Further geometric models of special importance in applications ranging from entertainment to ergonomic design and facial expression analysis are detailed 3D human face shapes. Thanks to advances in high-resolution acquisition systems in both space and time, there is an extensive amount of research in this area recently. As in the case of human bodies, defining a mathematical model describing possible shape variations of faces across different subjects and expressions directly is impractical, and generative models are commonly learned from databases of scans. The generative models that are employed differ from those reviewed in Part II because facial shape changes caused by pose are primarily driven by muscle movement rather than skeleton-based deformation models. Chapter 12 provides an overview of generative models used for 3D face modeling, in parts based on our previous review [26].

As in the case of human body shape, raw scans are typically registered and rigidly aligned prior to building generative models. In our work, we have developed a fully automatic template-based method to compute point-to-point correspondences between large databases of static human face models [100]. Unlike in the case of human body shapes, we model face deformations using blendshapes (see [71] for an overview). As in the case of body scans, we initialize the fitting procedure using a set of automatically predicted anatomic landmark positions on the face. The registered training data is subsequently used to learn statistical shape spaces. In our work, we started experimenting with global tensor-based shape spaces that allow to separate shape changes caused by identity and expression [16]. While we found these shape spaces to be of satisfactory quality overall, some artifacts were visible. Due to the "uncanny valley" problem in face modeling, these minor geometric artifacts can be highly visually noticeable.

The first visual artifact is caused by small drift in the computed correspondence information of the training data. We address this problem by jointly improving the initial correspondence and the resulting tensor-based

shape space using a groupwise approach. The developed method updates the initial correspondence across the entire training dataset, such that an energy is optimized that measures how compactly the resulting tensor-based model can be represented. This problem is challenging both in terms of its theoretical formulation and in terms of its implementation, as optimizing a groupwise energy function requires careful algorithmic choices and efficient coding strategies. This work was published in *International Conference on Computer Vision* [17], and Chapter 13 summarizes this work.

Tensor-based models require datasets for training that capture all combinations of all factors. For instance, all subjects of the training set need to be captured in every expression (among a fixed pre-chosen set) to learn models that capture identity and expression variation. This imposes strong limitations on learning the model, as data acquisition and processing is costly. To address this problem, we have recently extended our groupwise approach to allow learning high-quality generative models from incomplete training data that may in addition contain some training data that is corrupted by noise. This work was published in *Conference on Computer Vision and Pattern Recognition* [18], and Chapter 14 provides a summary.

The second visual artifact occurs because geometric detail is not well represented in global tensor-based shape spaces. To remedy this, we developed a multi-scale localized tensor-based shape space for human faces based on a wavelet decomposition of the training data. This work was published in *European Conference on Computer Vision* [25], and Chapter 15 summarizes this work.

For more technical details and experimental results on this part, the interested reader is referred to the original publications, which are available in electronic version at

[17]: `https://hal.inria.fr/hal-01205460/document`,

[18]: `https://hal.inria.fr/hal-01290783/document`, and

[25]: `https://arxiv.org/pdf/1401.2818.pdf`.

12 Statistical priors for 3D face modeling

Modeling and analyzing the 3D shape, motion and appearance of human faces has applications including entertainment, speech science, ergonomic design, the medical field, and security. For this reason, recent advances in high-resolution acquisition systems in both space and time have resulted in an extensive amount of research in 3D face shape and motion modeling within computer vision and graphics (e.g. [74, 117]). The following discussion focuses on *3D human face modeling* using *statistical deformation models* derived in a data-driven way. This chapter starts by giving a broad overview of works in this area, and subsequently defines the shape space used in our works.

Shape spaces for human faces

The first statistical shape model for 3D human faces is called *morphable model* and models both 3D shape and texture information of faces captured in a neutral expression using two independent PCA spaces [13]. The morphable model effectively represents 3D face shape using a simple representation, and has resulted in multiple follow-up works [4, 20, 36, 88]. The morphable model has been extended to handle expression variation by modeling offsets from the neutral expression using a separate PCA space [5]. This allows to model expression variation using a linear model.

Multiple works proposed to segment the face into localized regions and to learn independent morphable models for each part, thereby increasing the variability captured by the model. While most methods segment the face manually [10, 13, 61, 113], Smet and van Gool [110] find the segmentation automatically using clusters of vertices computed based on features derived from vertex displacements over the training set. Is is also possible

to capture and combine localized shape variations in different areas using a multi-resolution framework based on a wavelet decomposition of the training data [27]. Another strategy to add fine-scale facial detail to the model is based on hierarchical pyramids that model the difference between a smooth face and one with high-resolution detail such as wrinkles [50]. Localized patch-based models have also been proposed for recognition tasks such as expression recognition [79].

Multilinear models were introduced to allow for varying facial expressions [123]. These tensor-based models decouple different factors that impact the 3D face shape by modeling them as independent. These factors may include expression, viseme, and age. Multilinear models effectively model global 3D shape changes of faces and have been used to reconstruct sequences of 3D faces from video [31, 37] and extended to be trained from partial data in an unsupervised way [125].

Li et al. [74] proposed a generative model that captures fine-scale shape changes due to facial motion. In this model, shape changes due to skeletal motion are modeled using a skinning approach, while shape changes due to identity and expression are modeled as linear factors as in the morphable model.

A different line of works analyzes face shape without the need to compute point-to-point correspondences beforehand. Based on a single point match (the tip of the nose), corresponding geodesic isocontours are extracted and used for analysis [101].

Recently, deep learning techniques have started to influence the field of 3D face modeling (e.g. [93, 94, 118, 143]). In particular, some works have started to combine encoders based on convolutional neural networks with linear generative models used as decoder for the problem of 3D face reconstruction from 2D photo and video [69, 69, 105, 117]. In our work, we have recently extended this line of work to learn a generative 3D face model using an autoencoder architecture that enforces the structure of the decoder to that of a multilinear model [46].

Multilinear model

The works discussed in this part are based on the tensor-based multilinear model that was first applied to 3D face modeling by Vlasic et al. [123]. In this

work, we focus on multilinear models that represent two modes of variation, namely variation due to identity and due to pose (i.e. expression). Such a multilinear model represents the coordinate vector $\mathbf{c} \in \mathbb{R}^{3n}$ of a registered 3D face with two parameter vectors $\mathbf{w}_{id} \in \mathbb{R}^{d_{id}}$ and $\mathbf{w}_{pose} \in \mathbb{R}^{d_{pose}}$, and allows to approximate $\mathbf{c}$ using the generator function

$$\mathbf{c}\left(\mathbf{w}_{id}, \mathbf{w}_{pose}\right) = \mathcal{M} \times_2 \mathbf{w}_{id}^T \times_3 \mathbf{w}_{pose}^T + \bar{\mathbf{c}}, \tag{12.1}$$

where $\mathcal{M} \in \mathbb{R}^{3n \times d_{id} \times d_{pose}}$ is a 3-mode tensor called *core tensor* of the model, and $\times_2$ and $\times_3$ denote tensor-matrix multiplications along the second and third modes. Here, $\bar{\mathbf{c}}$ is the mean computed over all training data (all identities captured in all expressions) after rigid alignment. Note that this generator function has the form given in Equation 8.2.

The standard way to learn the core tensor $\mathcal{M}$ is to start with a registered aligned training set of faces $\mathbf{c}_i$, where each identity used for training was captured in the same set of expressions. The training data are then stacked into a tensor $\mathcal{A}$, such that the vertex coordinates are stacked along the first mode, identities along the second mode, and expressions along the third mode. The tensor $\mathcal{A}$ is then decomposed using a Tucker decomposition into a core tensor $\mathcal{M}$ and two orthogonal factor matrices. The goal is to find $\mathcal{M}$ of fixed dimensions d_{id} and d_{pose} (smaller than the number of identities and expressions present in the training data, respectively), while minimizing the norm of the residual between $\mathcal{A}$ and the reconstructed tensor from the reduced core tensor $\mathcal{M}$. This can be seen as an extension of PCA to the multilinear case. The main difference to PCA is that optimizing this problem is known to be NP-hard [56]. Furthermore, Tucker decompositions are not unique in general. Different methods have been proposed in the literature to computer Tucker decompositions (see e.g. [66] for an overview). After experimenting with different decompositions, we have chosen to use higher-over singular value decomposition (HOSVD) [40] to compute a Tucker decomposition in our works, as it is simple to compute, yet yields good approximations in our application.

The multilinear model can be combined with a statistical prior as follows. If we consider only a single mode, the multilinear model reduces to a PCA space. PCA fits a multivariate Gaussian distribution to the training data, and for data that can be approximated well this distribution is often used successfully as prior. For a multilinear model, we can fit two multivariate Gaussian distributions to the data: one in identity space and one in pose space.

13 Groupwise multilinear correspondence optimization for 3D faces

Multilinear models effectively represent face shape in a low-dimensional parameter space that decouples shape variation due to identity and expression. However, when building a multilinear model from template-fitted scan data, we found variance in the model that results in drift along the surface. This variance is due to inaccuracies in the point-to-point correspondence of the training data. To remedy this, starting from the (inaccurate) multilinear model as initialization, we jointly optimize both the multilinear model and the registration of the 3D scans used for training. Our method proceeds by measuring the quality of the multilinear model, and by changing the registration of the training data such that the quality of the multilinear model improves. Since the quality of the multilinear model depends on all shapes, this is a groupwise optimization method. The quality of the multilinear model is defined based on the minimum description length principle.

Overview

The basic idea of the minimum description length (MDL) principle is to minimize the length of a message that needs to be transmitted from a sender to a receiver in order to transmit a fixed amount of information. MDL has been applied to jointly optimize the registration and the resulting statistical model of a set of 3D shapes. The key idea is to solve a groupwise correspondence optimization problem that explicitly favors compact models (thereby allowing to encode each shape in a low-dimensional shape space, or using a short message). This idea has been applied most often to linear shape spaces learned using PCA [38, 49, 67], and our work extends these methods to the multilinear model case.

Our approach improves the initial multilinear model iteratively. Each iteration measures the quality of the multilinear model using a groupwise objective function, which is optimized in a continuous parameter space using a quasi-Newton method.

Groupwise objective function

The groupwise objective function we optimize contains a compactness energy $E_{compact}$, which aims to find the multilinear model of minimal dimensionality that can describe the training data, and a regularization energy E_{reg}, which ensures that all parts of the training shapes are encoded in shape space. That is, we minimize

$$E_{mdl} = E_{compact} + \omega_{reg} E_{reg}, \tag{13.1}$$

where the weight ω_{reg} controls the influence of E_{reg}. We measure $E_{compact}$ as the percentage of data variability represented by the first k components of each mode of the multilinar model. To encourage sparsity, we follow Kotcheff and Taylor [67] and choose a log-sum penalty function. We optimize E_{mdl} by moving vertices within a continuous surface representation of the training shapes. Therefore, $E_{compact}$ can be minimized by moving points away from regions that contain geometric detail with high variability, such as the region of the nose. To avoid undersampling in these regions, the regularization term E_{reg} is a discretized bi-Laplacian [65] that in our case, where points are constrained to move on the surface of the face, encourages points to be regularly distributed over each training shape.

Parameterization of training data

The energy E_{mdl} is optimized by moving the (registered) vertices along the continuous surfaces of the training shapes. To perform this optimization, we compute a continuous surface representation for each training shape as follows. First, we unwrap the 3D template mesh that was used to compute the initial correspondence into 2D parameter space. Since all faces are registered, this provides a discrete parametrization for all registered faces of the training data. The parameterization chosen in our work is visualized in Figure 13.1. Second, this discrete parameterization is used to define a continuous mapping from

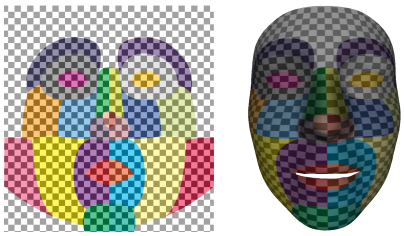

Figure 13.1: *Initial surface parametrization of the 3D face template. Left: 2D parameter domain. Right: 3D parametrization. Figure from [17].*

2D parameter space to the surface of each training face using a thin-plate spline mapping [19].

Optimization

The goal is to optimize E_{mdl} defined in Equation 13.1 with respect to the vertices in 2D parameter space of all training shapes. Thanks to the choice of a continuous surface representation, E_{mdl} is differentiable with respect to these parameters. However, optimizing over all parameters jointly is not feasible due to the large number of parameters. To remedy this, the optimization is performed over the parameters of each training shape individually. Each iteration considers all training shapes in a randomized order to avoid bias towards any particular shape.

An additional problem is that face models contain boundaries (for the mouth and an outer boundary in our case). This implies that E_{mdl} can be minimized by collapsing the entire surface into a single point. To prevent this solution, we enforce boundary conditions during the optimization.

We have shown that the proposed approach is significantly more efficient than existing MDL-based approaches that use a PCA-based statistical model. Figure 13.2 shows some qualitative optimization results, where the initial registrations were corrupted by random noise. The results are obtained after 15 iterations of our method. Note that the proposed approach is robust with respect to fairly large errors in the initialization.

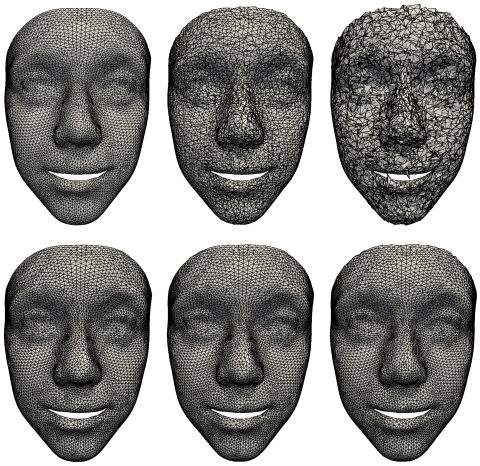

Figure 13.2: *Noise in the initialization before (top) and after (bottom) optimization. Left to right: no, low, and high noise. Figure from [17].*

14 Robust multilinear model learning

One problem when learning multilinear models is that they degrade when learned from training data where not each person was acquired in every expression, where some of the scans are noisy or partially occluded, or where some labels are erroneous. This imposes strong limitations on the training data, and unfortunately disqualifies many existing static 3D face datasets that have been captured because problems are present in a small percentage of the data. For example, the Bosphorus database [102] contains information on action units, but not every subject was scanned in every action unit. This chapter presents a new method to leverage such data for training a multilinear model by again pursuing a groupwise approach that jointly learns the model and fixes the training data.

The reason for jointly learning the model and fixing the training data is that computing the multilinear model is a chicken-and-egg problem. Given a complete set of training data showing each person in every expression allows to compute a multilinear model of high quality. Given such a model, it is possible to complete missing data, reconstruct occluded or noisy areas in scans, and recognize expressions (see e.g. [30, 83]). The joint optimization is achieved by extending the work presented in Chapter 13 by modifying the groupwise objective function.

Groupwise objective function

To handle missing and corrupt data, the groupwise energy function of Equation 13.1 is extended as follows

$$E_{robust} = E_{compact} + \omega_{reg} E_{reg} + \omega_{data} E_{data}, \tag{14.1}$$

where $E_{compact}$ is the MDL-based multilinear compactness term as before, E_{reg} is a regularization energy, E_{data} is a data term, and ω_{reg} and ω_{data} are

weights that control the influence of the different energy terms. Unlike in Equation 13.1, E_{reg} uses a bi-Laplacian energy to measure the difference between a coordinate vector $\mathbf{c}$ and a reference shape $\tilde{\mathbf{c}}$. This allows to regularize $\mathbf{c}$ by asking it to stay close to $\tilde{\mathbf{c}}$. This stronger regularization is required as E_{robust} is optimized with respect to the 3D vertex positions of the missing or corrupt training shapes $\mathbf{c}$, which leads to more degrees of freedom during the optimization than in Chapter 13, where the optimization is performed with respect to points on the known surface of the face scan. In addition to these terms, the data term E_{data} is used for training scans known to be plagued by occlusion or acquisition noise. The energy E_{data} is a sum of truncated squared distances between vertices on $\mathbf{c}$ and their nearest neighbors on the corresponding scan.

The input to the algorithm is a set of at most $d_{id} \cdot d_{pose}$ registered shapes $\mathbf{c}_{i,p}$ with $i \in \{1, \cdots, d_{id}\}$ and $p \in \{1, \cdots, d_{pose}\}$. Furthermore, for the remaining shapes, partial and possibly noisy scans $\mathbf{s}_{i,p}$ may be provided. The data $\mathbf{s}_{i,p}$ are not in correspondence with the registered shapes, but are required to be rigidly aligned to the registered shapes $\mathbf{c}_{i,p}$. The provided labels i, p are used as initial semantic labeling of the data. Label pairs for which no information is provided are called missing shapes, and we will estimate these shapes using Equation 14.1.

Optimization

Each missing or corrupt shape is initialized by the mean over the registered available training shapes of the same identity i and pose p. This results in a full tensor of (possibly unreliable) training data. The algorithm starts by optimizing the semantic labeling, and subsequently optimizes E_{robust} for each corrupt or missing shape individually.

To optimize semantic labeling, the labels of the data are permuted by reorganizing coordinate vectors $\mathbf{c}$ in the training tensor. As the coordinate vectors $\mathbf{c}$ are left unchanged during this optimization, we set $\omega_{reg} = \omega_{data} = 0$ in Equation 14.1. A joint optimization of the semantic labels over all data is infeasible. Furthermore, we observed no errors in identity labels in our training data. Hence, we optimize E_{robust} for each identity individually and iterate over all identities in random order. For each identity i we aim to find the permutation $\pi_i = \{\pi_i(1), \ldots, \pi_i(d_{pose})\}$ with $\pi_i(p) \in \{1, \ldots, d_{pose}\}$ of the available poses that minimizes E_{robust}. We solve this integer problem using a

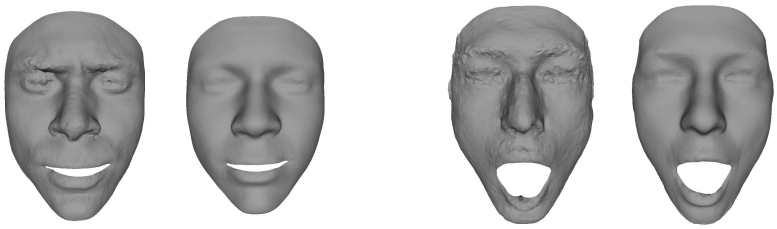

Figure 14.1: *Robustness to missing data. Each pair shows the ground truth on the left and our result on the right. Figure adapted from [18].*

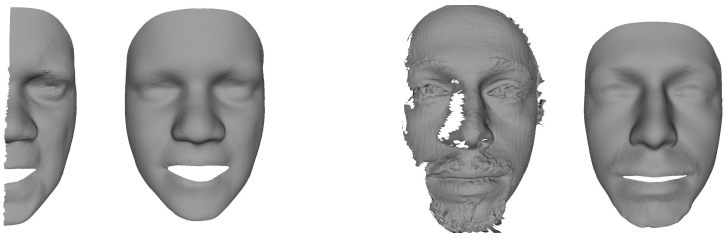

Figure 14.2: *Robustness to corrupt data. Each pair shows the partial data on the left and our result on the right. Figure adapted from [18].*

threshold accepting method [85].

For the remaining optimization, the algorithm distinguishes missing shapes and corrupted shapes. To estimate a missing shape, we set $\omega_{data} = 0$ in Equation 14.1 as no input scan data is provided. To estimate a shape for which a partial or noisy scan $\mathbf{s}_{i,p}$ is available, all energy terms in Equation 14.1 are weighted by positive weights. The energy E_{robust} is then optimized with respect to the coordinate vector $\mathbf{c}_{i,p}$ that contains the 3D coordinates of the missing shape. For both missing a corrupted shapes, the reference shape in E_{reg} is set to the initial shape computed by averaging the available training shapes of the same identity and pose. In our implementation, we optimize E_{robust} using a quasi-Newton method.

Figure 14.1 shows a qualitative result obtained by leaving 10% of the available data out of the training set and by estimating these shapes. Note that the estimates are geometrically close to the ground truth. Figure 14.2 shows a similar qualitative result where 10% of the data were chosen to have missing parts or geometric noise.

15 Multilinear wavelet model

The multilinear models discussed in this part allow to effectively represent the overall global shape of a 3D face for varying identities and expressions in a low-dimensional shape space. However, these global statistical models do not capture fine-scale detail. As geometric detail is often required to model realistic 3D faces, we have extended this global multilinear model to capture more geometric detail by decomposing the surface of the face using a wavelet transform, and by learning localized multilinear models on the resulting coefficients.

The key idea of the multilinar wavelet model is to decompose the registered training faces using a spherical subdivision wavelet, thereby generating a set of corresponding wavelet coefficients. While the vertex positions of the training faces are highly correlated, the wavelet coefficients are decorrelated. This implies that we can treat the coefficients separately, and we take advantage of this by learning distinct multilinear models per coefficient. Figure 15.1 illustrates this idea.

The spherical wavelets we use operate on a subdivision surface (following a Catmull-Clark subdivision scheme) [12]. This allows to represent coarse-scale shape properties by a few coefficients, while representing localized fine-scale details by additional coefficients.

We devised an algorithm to fit this multilinear wavelet shape space to input scans that are possibly corrupted by noise as long as a few landmarks are available to guide the fitting. Figure 15.2 shows some fitting results, and in particular compares the results obtained using a multilinear wavelet model with the results obtained using a global multilinear model. Note that the multilinear wavelet model allows to capture more geometric detail.

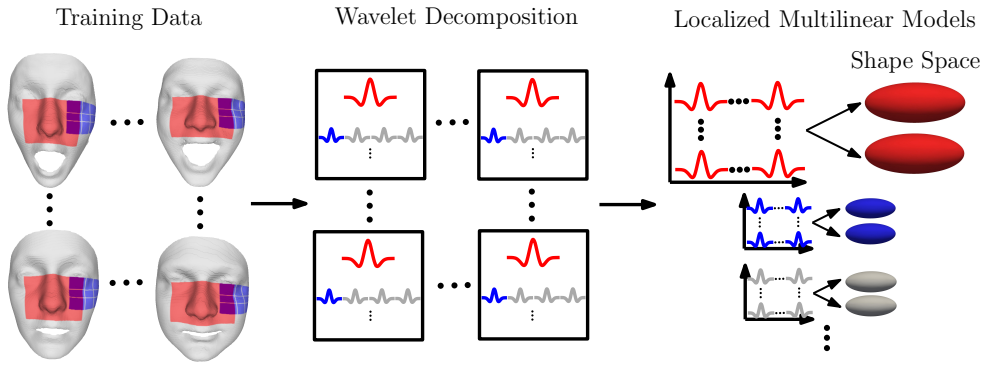

Figure 15.1: *Overview of the training. Left: Training data with highlighted impact of the basis function. Middle: Wavelet decomposition of each face of the training data. Right: Corresponding wavelet coeefficients and learned multilinear model shape spaces. Figure from [25].*

Advantages and disadvantages

The global model presented in Equation 12.1 is much more commonly used than localized models similar in spirit to the one presented in this chapter. The main advantages of the global model over such localized models is its compact representation. In our works, we typically set $d_{id} \leq 100$ and $d_{pose} \leq 10$ for the global multilinear model. The wavelet model presented in this chapter represents each wavelet coefficient using 6 dimensions. That is, if we start from a 5×7 quad mesh as a base mesh, the coarsest level is represented using $5 \cdot 7 \cdot 6 = 210$ coefficients in shape space. The subdivision of each cell will generate 4 additional cells, and the coefficients corresponding to each cell are again represented using 6 dimensions in shape space. This leads to high-dimensional shape spaces.

The key advantage of the global multilinear model is that it allows to visualize and intuitively understand the main modes of variation. This allows to study the geometric variations of the rough overall face shape.

The key advantage of the local multilinear model is that more general shape spaces are obtained as more detail is encoded. This implies directly that such detail can be recovered during fitting applications, where we optimize for points in shape space that fit to (possibly noisy or partial) input scans. A further advantage is that this optimization can be solved faster than for global multilinear models, as multiple low-dimensional optimization problems are solved (6-dimensional problems in our case) rather than one high-dimensional

Figure 15.2: *Each row shows a reconstruction example for noisy scans in different expressions. From left to right: input data, fitting a multilinear model, fitting a multilinear wavelet model. Figure adapted from [25].*

one (of dimension $d_{id} + d_{pose}$ in case of global multilinear models).

In summary, whether a global or local shape space is more suitable is application dependent. Geometric detail can be important in applications including ergonomic design, tele-presence, and high-quality facial capture. Less detail and a lower-dimensional model are typically advantageous for applications including recognition tasks (face and expression recognition), model-based image and video modification, and the control of virtual avatars.

16 Conclusions and future work

This chapter presents a short summary and concluding remarks concerning the contributions presented in this document. Furthermore, we discuss some interesting directions for future research.

Concluding remarks

The work presented in this manuscript concerns the processing and analysis of 3D human shapes, and discusses three classes of deformation models that have been used in our work to analyze the deformations of clothing and loco-motion of humans, human body shape, and human face shape.

The first model was developed for shapes that are near-isometric while their 3D acquisitions are far from near-isometric in local regions due to acquisition artifacts caused by occlusion and scanner noise. This model allows to find partial isometries using a low-dimensional search space, and to combine these parts using a clustering approach. Using this approach allows for robust shape processing as discrepancies in the intrinsic geometry of the two shapes are found as boundaries of matching parts. Using this model allows extensions to other applications. We have demonstrated this by using the model to detect partial intrinsic symmetries and to spatially align motion sequences of dressed humans.

The second model represents human body shapes using a parameter space that decouples the influence of parameters that control shape changes caused by identity and posture. The chosen model allows for efficient reconstructions of 3D shapes from the representation. We have shown how to effectively build such a model using a large training database of raw 3D scans by processing these scans automatically using best-practice solutions. The resulting shape space can aid in solving a number of under-constrained problems such as

reconstruction problems from partial data (including 2D images and partial scans). We have demonstrated this point by using the model to estimate body shape and motion under wide clothing.

The third class of models represent human face shapes using parameter spaces that decouple the influene of parameters that control shape changes caused by identity and expression. The first model we discussed is a compact global model that can be trained from a complete training tensor of registered shapes. We have presented a method to train this model while correcting errors in the provided initial registration using a groupwise approach, which in combination with an automatic template-fitting method to compute initial registrations allows to build such models fully automatically from new raw training datasets. As a further extension, we extended this groupwise approach to allow for small amounts of missing data and scans corrupted by significant noise in the training data. This allows to leverage existing databases of 3D face scans to train more expressive models. Finally, we presented a contribution that extends this class of models to represent geometric information at multiple scales, thereby allowing to represent fine-scale geometric detail. This is important for face modeling as human observers are very good at recognizing overly smooth 3D face shape.

While studying the problems presented in this manuscript, we have learned two important lessons.

Implementation choices are important As expected, the deformation model chosen to represent the desired deformations in a constrained shape space has a significant influence on the complexity of the resulting optimization problems. However, the deformation model is only part of the choice that determines whether problems can be solved efficiently. In practice, we found it equally important to consider different implementation choices for the optimization problems. The reason is that the practical value of a model is limited if we cannot optimize for the unknown parameters efficiently and in a numerically stable way. One example of this occurs in the work presented in Chapter 13. If we aim to optimize for the 3D vertex positions of every face, further regularizers need to be considered that aim to constrain the deviation from the initial face shape. By optimizing in a continuous 2D parameter space, the optimization problem not only requires less energy terms, but also becomes lower dimensional. The choice of algorithmic tool used for the optimization can also make a significant difference in both the effectiveness and the efficiency of the algorithm.

Validation needs to consider application We found that different deformation models are often used at the same time for different applications. The reason is that there are a number of trade-offs. One example is the trade-off between the computational time required to reconstruct a shape from parameter space and the ability of the space to accurately represent the shape variability present in the training data (as discussed in Part II). Another one is the trade-off between the ability of a shape space to represent 3D geometric detail and the compactness of the representation (as discussed in Part III). This implies that experimental validation should consider the applications of interest or provide an analysis on which applications might benefit the most from the presented model.

Future work

We conclude this manuscript with some long-term goals for future work in this area.

Combining more and less detailed models

One interesting avenue for future work is to integrate the deformation models developed for different body parts or accessories into one global model allowing to represent the detailed geometry and deformation behaviour of humans and their accessories. For instance, the multilinear face models (see Part III) could ideally replace the static and geometrically coarse face information of the S-SCAPE body models (see Part II), while near-isometrically deforming clothes (see Part I) could form a layer on top of the body.

Developing such models is challenging. As the geometry and motion of the full body and its parts are correlated, a combined model needs to be trained from data that contains a global view of the body, yet enough resolution in localized regions (such as the face) to allow for model building. Existing datasets that capture the full body lack such fine-scale detail in localized regions. Another challenge when building combined models is to find a good trade-off between representing geometric detail and the dimensionality of the combined parameter space. One the one hand, representing high levels of geometric detail may easily lead to high-dimensional parameter spaces that are impractical to use in optimization problems that typically occur when fitting the model to partial data. On the other hand, restricting the space to

be low-dimensional may lead to problems of lacking geometric detail that are easily spotted by human observers.

A recent work combines parametric models for full body and hand deformations [98].

Combining generative and physical models

The works presented in this manuscript do not rely on physical modeling. Instead, generative models are learned based on training datasets of human body and face shape, while cloth is modeled using near-isometric deformations. Combining generative and physical models is a promising research direction, and some recent works explore this avenue for human bodies [63].

This combination is particularly interesting to model clothing because it is challenging to model complex clothing deformations using physical simulation alone, while it is challenging to capture enough training data to build expressive generative models for clothing. One recent technique in this direction allows to recover physical material parameters of cloth based on 2D video information [137].

Statistical analysis of motion data

Over the past two decades, two directions of statistical analysis have progressed significantly. On the one hand, statistical models of 3D human body and face shape learned from a database of static dense 3D scans have been applied to problems in computer vision and graphics. Recent work enhance such statistical models to include dynamic shape changes (e.g. [90]). While these models capture fine-scale geometric detail, they does not capture the structure of the motion. On the other hand, statistical models of 3D body motion learned from trajectories of sparse 3D markers have been applied to solve computer vision problems (e.g. [41, 119]). While these statistical models capture sparse large-scale motion patterns, they lack dense geometric information. Combining these two powerful concepts is promising. However, statistical models that aim to capture both detailed geometric information and motion patterns jointly have received very little study to date. One notable exception combines a dense model of static body shape with a sparse model of motion [68]. While this is a promising first step, the computation of the model relies heavily on the annotation of the motion data and does not

model dynamics of the motion. One avenue for future work in this direction that is worth exploring is to leverage techniques based on deep learning that have recently begun to influence the field of 3D shape modeling to build generative models of 3D shape in motion.

Bibliography

[1] B. Allain, J.-S. Franco, and E. Boyer. An efficient volumetric framework for shape tracking. In *Conference on Computer Vision and Pattern Recognition*, 2015.

[2] B. Allain, J.-S. Franco, E. Boyer, and T. Tung. On mean pose and variability of 3d deformable models. In *European Conference on Computer Vision*, 2014.

[3] 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.

[4] B. Amberg, A. Blake, A. Fitzgibbon, S. Romdhani, and T. Vetter. Reconstructing high quality face-surfaces using model based stereo. In *International Conference on Computer Vision*, 2007.

[5] B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3D face recognition with a morphable model. In *International Conference on Automatic Face Gesture Recognition*, 2008.

[6] 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.

[7] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *ICCV Workshop on Dynamic Shape Capture and Analysis*, 2011.

[8] A. O. Balan and M. J. Black. The naked truth: Estimating body shape under clothing. In *European Conference on Computer Vision*, 2008.

[9] A. Bas, W. A. P. Smith, T. Bolkart, and S. Wuhrer. Fitting a 3d morphable model to edges: A comparison between hard and soft correspondences. In *ACCV Workshop on Facial Informatics*, 2016.

[10] C. Basso and A. Verri. Fitting 3d morphable models using implicit representations. In *Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2007.

[11] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer, 2002.

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

[13] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Conference on Computer graphics and interactive techniques*, 1999.

[14] 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.

[15] T. Bolkart, P. Bose, C. Shu, and S. Wuhrer. A general framework to generate sizing systems from 3d motion data applied to face mask design. In *3D Vision*, 2014.

[16] T. Bolkart and S. Wuhrer. 3d faces in motion: Fully automatic registration and statistical analysis. *Computer Vision and Image Understanding*, 131:100–115, 2015.

[17] T. Bolkart and S. Wuhrer. A groupwise multilinear correspondence optimization for 3d faces. In *International Conference on Computer Vision*, 2015.

[18] T. Bolkart and S. Wuhrer. A robust multilinear model learning framework for 3d faces. In *Conference on Computer Vision and Pattern Recognition*, 2016.

[19] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *Transactions on Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.

[20] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway. A 3d morphable model learnt from 10,000 faces. In *Conference on Computer Vision and Pattern Recognition*, 2016.

[21] A. Bronstein, M. Bronstein, A. Bruckstein, and R. Kimmel. Partial similarity of objects, or how to compare a centaur to a horse. *International Journal on Computer Vision*, 84(2):163–183, 2009.

[22] A. Bronstein, M. Bronstein, and R. Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer, 2008.

[23] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Robust expression-invariant face recognition from partially missing data. In *European Conference on Computer Vision*, 2006.

[24] M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Conference on Computer Vision and Pattern Recognition*, 2010.

[25] A. Brunton, T. Bolkart, and S. Wuhrer. Multilinear wavelets: A statistical shape space for human faces. In *European Conference on Computer Vision*, 2014.

[26] 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*, 128:1–17, 2014.

[27] 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*, 2011.

[28] A. Brunton, M. Wand, S. Wuhrer, H.-P. Seidel, and T. Weinkauf. A low-dimensional representation for robust partial isometric correspondences computation. *Graphical Models*, 76:70–85, 2014.

[29] C. Budd, P. Huang, M. Klaudiny, and A. Hilton. Global non-rigid alignment of surface sequences. *International Journal of Computer Vision*, 102:256–270, 2013.

[30] C. Cao, Y. Weng, S. Lin, and K. Zhou. 3d shape regression for real-time facial animation. *ACM Transactions on Graphics*, 32(4):41:1–10, 2013.

[31] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. FaceWarehouse: a 3D Facial Expression Database for Visual Computing. *Transactions on Visualization and Computer Graphics*, 3:413–425, 2014.

[32] W. Chang, H. Li, N. Mitra, M. Pauly, and M. Wand. Dynamic geometry processing. In *Eurographics tutorial*, 2012.

[33] Y. Chen and R. Cipolla. Learning shape priors for single view reconstruction. In *Workshop on 3D Imaging and Modelling*, 2009.

[34] Y. Chen, Z. Liu, and Z. Zhang. Tensor-based human body modeling. In *Conference on Computer Vision and Pattern Recognition*, 2013.

[35] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics*, 34(4):#69:1–13, 2015.

[36] H. Dai, N. Pears, W. Smith, and C. Duncan. A 3d morphable model of craniofacial shape and texture variation. In *International Conference on Computer Vision*, 2017.

[37] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister. Video face replacement. *ACM Transactions on Graphics*, 30(6):130:1–10, 2011.

[38] R. Davies, C. Twining, and C. Taylor. *Statistical Models of Shape: Optimisation and Evaluation*. Springer, 2008.

[39] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. *ACM Transactions on Graphics*, 27(3):#98,1–10, 2008.

[40] L. De Lathauwer. *Signal processing based on multilinear algebra*. PhD thesis, K.U. Leuven, Belgium, 1997.

[41] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. D. Bimbo. 3d human action recognition by shape analysis of motion trajectories on riemannian manifold. *Transactions on Cybernetics*, 45(7):1340–1352, 2015.

[42] E. Dibra, H. Jain, C. Oztireli, R. Ziegler, and M. Gross. Hs-nets : Estimating human body shape from silhouettes with convolutional neural networks. In *3DV*, 2016.

[43] M. Dou, S. Khamis, Y. Degtyarev, P. Davidson, S. Fanello, A. Kowdle, S. Escolano, C. Rhemann, D. Kim, J. Taylor, P. Kohli, V. Tankovich, and S. Izadi. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics*, 35(4):#114:1–13, 2016.

[44] I. Dryden and K. Mardia. *Statistical Shape Analysis*. Wiley, 2002.

[45] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1285–1295, 2003.

[46] V. Fernàndez Abrevaya, S. Wuhrer, and E. Boyer. Multilinear autoencoder for 3d face model learning. In *Winter Conference on Applications of Computer Vision*, 2018.

[47] N. Gelfand and L. Giubas. Shape segmentation using local slippage analysis. In *Symposium on Geometry Processing*, 2004.

[48] P. Ghosh, J. Song, E. Aksan, and O. Hilliges. Learning human motion models for long-term predictions. In *3DV*, 2017.

[49] S. Gollmer, M. Kirschner, T. Buzug, and S. Wesarg. Using image semgmentation for evaluating 3d statistical shape models built with groupwise correspondence optimization. *Computer Vision and Image Understanding*, 125:283–303, 2014.

[50] 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.

[51] P. Guan, L. Reiss, D. A. Hirshberg, A. Weiss, and M. J. Black. Drape: Dressing any person. *ACM Transactions on Graphics*, 31(4):35:1–10, 2012.

[52] P. Guan, A. Weiss, A. O. Bălan, and M. J. Black. Estimating human shape and pose from a single image. In *International Conference on Computer Vision*, 2009.

[53] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. In *Conference on Computer Vision and Pattern Recognition*, 2010.

[54] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. *Computer Graphics Forum*, 2(28):337346, 2009.

[55] A. Hewer, I. Steiner, and S. Wuhrer. A hybrid approach to 3d tongue modeling from vocal tract mri using unsupervised image segmentation and mesh deformation. In *Interspeech*, 2014.

[56] C. J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *Journal of the ACM*, 60(6):45:1–45:39, 2013.

[57] D. Hirshberg, M. Loper, E. Rachlin, and M. Black. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In *European Conference on Computer Vision*, 2012.

[58] Q. Huang, B. Adams, M. Wicke, and L. J. Guibas. Non-rigid registration under isometric deformations. *Computer Graphics Forum*, 27(5):1449–1457, 2008.

[59] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt. MovieReshape: tracking and reshaping of humans in videos. *ACM Transactions on Graphics*, 29:148:1–10, 2010.

[60] V. Jain, H. Zhang, and O. van Kaick. Non-rigid spectral correspondence of triangle meshes. *International Journal of Shape Modeling*, 13(1):101–124, 2007.

[61] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, L. Yunliang, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

[62] M. Kilian, N. J. Mitra, and H. Pottmann. Geometric modeling in shape space. *ACM Transactions on Graphics*, 26(3):64:1–8, 2007.

[63] M. Kim, G. Pons-Moll, S. Pujades, S. Bang, J. Kim, M. Black, and S.-H. Lee. Data-driven physics for soft-tissue animation. *ACM Transactions on Graphics*, 36(4):#54:1–12, 2017.

[64] V. G. Kim, Y. Lipman, and T. Funkhouser. Blended intrinsic maps. *ACM Transactions on Graphics*, 30(4):79:1–79:12, 2011.

[65] L. Kobbelt, S. Campagna, J. Vorsatz, and H.-P. Seidel. Interactive multi-resolution modeling on arbitrary meshes. In *Conference on Computer graphics and interactive techniques*, 1998.

[66] T. Kolda and B. Bader. Tensor decomposition and applications. *SIAM Review*, 52(3):455–500, 2009.

[67] A. C. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by direct optimization. *Medical Image Analysis*, 2(4):303 – 314, 1998.

[68] A. Kuznetsova, N. F. Troje, and B. Rosenhahn. A statistical model for coupled human shape and motion synthesis. In *International Conference on Computer Graphics Theory and Applications*, 2013.

[69] S. Laine, T. Karras, T. Aila, A. Herva, S. Saito, R. Yu, H. Li, and J. Lehtinen. Production-level facial performance capture using deep convolutional neural networks. In *Symposium on Computer Animation*, 2017.

[70] V. Leroy, J.-S. Franco, and E. Boyer. Multi-view dynamic shape refinement using local temporal integration. In *International Conference on Computer Vision*, 2017.

[71] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. Pighin, and Z. Deng. Practice and theory of blendshape facial models. In *Eurographics STAR report*, 2014.

[72] G. Li, X.-Z. Ye, and S.-Y. Zhang. An algorithm for filling complex holes in reverse engineering. *Engineering with Computers*, 24(2):119–125, 2008.

[73] H. Li, B. Adams, L. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics*, 28(5):175:1–175:10, 2009.

[74] T. Li, T. Bolkart, M. Black, H. Li, and J. Romero. Learning a model of facial shape and expression from 4d scans. *ACM Transactions on Graphics*, 36(6):#194:1–17, 2017.

[75] Y. Lipman, X. Chen, I. Daubechies, and T. A. Funkhouser. Symmetry factored embedding and distance. *ACM Transactions on Graphics*, 29(4):103:1–12, 2010.

[76] Y. Lipman and T. Funkhouser. Möbius voting for surface correspondence. *ACM Transactions on Graphics*, 28(3):72:1–72:12, 2009.

[77] Y. Lipman, R. M. Rustamov, and T. A. Funkhouser. Biharmonic distance. *ACM Transactions on Graphics*, 29(3):27:1–27:11, 2010.

[78] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. Black. SMPL: A Skinned Multi-Person Linear Model. *ACM Transactions on Graphics*, 34(6):248:1–248:16, 2015.

[79] A. Maalej, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berrett. Shape analysis of local facial patches for 3d facial expression recognition. *Pattern Recognition*, 44(8):1581–1589, 2011.

[80] E. L. Malvaer and M. Reimers. Geodesic polar coordinates on polugonal meshes. *Computer Graphics Forum*, 31(8):2423–2435, 2012.

[81] M. Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez, and M. Desbrun. Variance-Minimizing Transport Plans for Inter-surface Mapping. *ACM Transactions on Graphics*, 36(4):14:1–14, 2017.

[82] N. Mitra, M. Pauly, M. Wand, and D. Ceylan. Symmetry in 3d geometry: Extraction and applications. In *Eurographics State of the Art Report*, 2012.

[83] I. Mpiperis, S. Malassiotis, and M. G. Strintzis. Bilinear models for 3-D face and facial expression recognition. *Transactions on Information Forensics and Security*, 3:498–511, 2008.

[84] A. Neophytou and A. Hilton. Shape and pose space deformation for subject specific animation. In *Conference on 3D Vision*, 2013.

[85] V. Nissen and H. Paul. A modification of threshold accepting and its application to the quadratic assignment problem. *OR Spektrum*, 17(2-3):205–210, 1995.

[86] M. Ovsjanikov, Q. Mérigot, F. Mémoli, and L. J. Guibas. One point isometric matching with the heat kernel. *Computer Graphics Forum*, 29(5):1555–1564, 2010.

[87] R. Palais. On the differentiability of isometries. *Proceedings of the AMS*, 8(4):805–807, 1957.

[88] A. Patel and W. Smith. 3D morphable face models revisited. In *Conference on Computer Vision and Pattern Recognition*, 2009.

[89] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele. Building statistical shape spaces for 3d human modeling. *Pattern Recognition*, 67:276–286, 2017.

[90] G. Pons-Moll, J. Romero, N. Mahmood, and M. Black. Dyna: a model of dynamic human shape in motion. *ACM Transactions on Graphics*, 34(4):#120:1–14, 2015.

[91] D. Raviv, A. Bronstein, M. Bronstein, and R. Kimmel. Full and partial symmetries of non-rigid shapes. *International Journal of Computer Vision*, 89:18–39, 2010.

[92] M. Reuter, F.-E. Wolter, and N. Peinecke. Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. *Computer-Aided Design*, 38(4):342–366, 2006.

[93] E. Richardson, M. Sela, and R. Kimmel. 3d face reconstruction by learning from synthetic data. In *Conference on 3D Vision*, 2016.

[94] E. Richardson, M. Sela, R. Or-El, and R. Kimmel. Learning detailed face reconstruction from a single image. *Conference on Computer Vision and Pattern Recognition*, 2017.

[95] W. Rinow. Über Flächen mit Verschiebungselementen. *Mathematische Annalen*, 107:95–112, 1932.

[96] K. Robinette, H. Daanen, and E. Paquet. The CAESAR project: A 3-D surface anthropometry survey. In *Conference on 3D Digital Imaging and Modeling*, 1999.

[97] D. Rohmer, T. Popa, M.-P. Cani, S. Hahmann, and A. Sheffer. Animation wrinkling: Augmenting coarse cloth simulations with realistic-looking wrinkles. *ACM Transactions on Graphics*, 29(6):157:1–157:8, 2010.

[98] J. Romero, D. Tzionas, and M. J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics*, 36(6):#245:1–17, 2017.

[99] Y. Sahillioglu and Y. Yemez. Scale normalization for isometric shape matching. *Computer Graphics Forum*, 31(7):2233–2240, 2012.

[100] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto. Fully automatic expression-invariant face correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014.

[101] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen. An intrinsic framework for analysis of facial surfaces. *International Journal of Computer Vision*, 82:80–95, 2009.

[102] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *European Workshop on Biometrics and Identity Management*, 2008.

[103] R. Schmidt. Stroke parametrization. *Computer Graphics Forum*, 32(2):255–263, 2013.

[104] R. Schmidt, C. Grimm, and B. Wyvill. Interactive decal compositing with discrete exponential maps. *ACM Transactions on Graphics*, 25(3):605–613, 2006.

[105] M. Sela, E. Richardson, and R. Kimmel. Unrestricted facial geometry reconstruction using image-to-image translation. *International Conference on Computer Vision*, pages 1576–1585, 2017.

[106] H. Seo, Y. I. Yeo, and K. Wohn. 3D body reconstruction from photos based on range scan. In *Technologies for E-Learning and Digital Entertainment*, 2006.

[107] A. Sharma, R. P. Horaud, J. Cech, and E. Boyer. Topologically-robust 3D shape matching based on diffusion geometry and seed growing. In *Conference on Computer Vision and Pattern Recognition*, 2011.

[108] A. Shehu, A. Brunton, S. Wuhrer, and M. Wand. Characterization of partial intrinsic symmetries. In *ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, 2014.

[109] A. Shehu, J. Yang, J.-S. Franco, F. Hétroy-Wheeler, and S. Wuhrer. Computing temporal alignments of human motion sequences in wide clothing using geodesic patches. In *Conference on 3D Vision*, 2016.

[110] M. Smet and L. V. Gool. Optimal regions for linear model-based 3d face reconstruction. In *Asian Conference on Computer Vision*, 2010.

[111] C. Stoll, J. Gall, E. de Aguiar, S. Thrun, and C. Theobalt. Video-based reconstruction of animatable human characters. *ACM Transactions on Graphics*, 29:139:1–10, 2010.

[112] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature-based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

[113] F. ter Haar and R. Veltkamp. 3d face model fitting for recognition. In *European Conference on Computer Vision*, 2008.

[114] A. Tevs, A. Berner, M. Wand, I. Ihrke, M. Bokeloh, J. Kerber, and H.-P. Seidel. Animation cartography - intrinsic reconstruction of shape and motion. *ACM Transactions on Graphics*, 31(2):#12:1–15, 2012.

[115] A. Tevs, A. Berner, M. Wand, I. Ihrke, and H.-P. Seidel. Intrinsic shape matching by planned landmark sampling. *Comuputer Graphics Forum*, 29(2):543–552, 2011.

[116] A. Tevs, M. Bokeloh, M. Wand, A. Schilling, and H.-P. Seidel. Isometric registration of ambiguous and partial data. In *Conference on Computer Vision and Pattern Recognition*, 2009.

[117] A. Tewari, M. Zollhöfer, H. Kin, P. G. G. Bernard, P. Pérez, and C. Theobalt. MoFa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *International Conference on Computer Vision*, 2017.

[118] A. T. Tran, T. Hassner, I. Masi, and G. Medioni. Regressing robust and discriminative 3d morphable models with a very deep neural network. *Conference on Computer Vision and Pattern Recognition*, 2017.

[119] N. Troje. Retrieving information from human movement patterns. In T. Shipley and J. Zacks, editors, *Understanding Events: From Perception to Action*, pages 308–334. Oxford University Press, 2008.

[120] A. Tsoli, N. Mahmood, and M. Black. Breathing life into shape: Capturing, modeling and animating 3d human breathing. *ACM Transactions on Graphics*, 33:1–11, 2014.

[121] O. van Kaick, H. Zhang, and G. Hamarneh. Bilateral maps for partial matching. *Computer Graphics Forum*, 32(6):189–200, 2013.

[122] D. Vlasic, I. Baran, W. Matusik, and J. Popović. Articulated mesh animation from multi-view silhouettes. *ACM Transactions on Graphics*, 27(3):97, 2008.

[123] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. *ACM Transactions on Graphics*, 24(3):426–433, 2005.

[124] M. Wand, B. Adams, M. Ovsjanikov, A. Berner, M. Bokeloh, P. Jenke, L. Guibas, H.-P. Seidel, and A. Schilling. Efficient reconstruction of nonrigid shape and motion from real-time 3D scanner data. *ACM Transactions on Graphics*, 28(2):1–15, 2009.

[125] M. Wang, Y. Panagakis, P. Snape, and StefanosZafeiriou. Learning the multilinear structure of visual data. In *Conference on Computer Vision and Pattern Recognition*, 2017.

[126] A. Weiss, D. Hirshberg, and M. Black. Home 3D body scans from noisy image and range data. In *International Conference on Computer Vision*, 2011.

[127] S. Wuhrer, Z. Ben Azouz, and Shu. Posture invariant surface description and feature extraction. In *Conference on Computer Vision and Pattern Recognition*, 2010.

[128] S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, and J. Lang. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding*, 127:31–42, 2014.

[129] S. Wuhrer and C. Shu. Shape from suggestive contours using 3d priors. In *Canadian Conference on Computer and Robot Vision*, 2012.

[130] S. Wuhrer and C. Shu. Estimating 3d human shapes from measurements. *Machine Vision and Applications*, 24(6):1133–1147, 2013.

[131] S. Wuhrer, C. Shu, Z. B. Azouz, and P. Bose. Posture invariant correspondence of incomplete triangular manifolds. *International Journal of Shape Modeling*, 13(2):139–157, 2007.

[132] S. Wuhrer, C. Shu, and P. Xi. Landmark-free posture invariant human shape correspondence. *The Visual Computer*, 27(9):843–852, 2011.

[133] S. Wuhrer, C. Shu, and P. Xi. Posture-invariant statistical shape analysis using Laplace operator. *Computers & Graphics*, 36(5):410–416, 2012.

[134] P. Xi, W.-S. Lee, and C. Shu. A data-driven approach to human-body cloning using a segmented body database. In *Pacific Conference on Computer Graphics and Applications*, 2007.

[135] K. Xu, H. Zhang, W. Jiang, R. Dyer, Z. Cheng, L. Liu, and B. Chen. Multi-scale partial intrinsic symmetry detection. *ACM Transactions on Graphics*, 31(6):#181:1–12, 2012.

[136] J. Yang, J.-S. Franco, F. Hétroy-Wheeler, and S. Wuhrer. Estimation of human body shape in motion with wide clothing. In *European Conference on Computer Vision*, 2016.

[137] S. Yang, J. Liang, and M. Lin. Learning-based cloth material recovery from video. In *International Conference on Computer Vision*, 2017.

[138] Y. Yang, D. Günther, S. Wuhrer, A. Brunton, I. Ivrissimtzis, H.-P. Seidel, and T. Weinkauf. Correspondences of persistent feature points on near-isometric surfaces. In *ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, 2012.

[139] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3D dynamic facial expression database. In *International Conference on Automatic Face & Gesture Recognition*, pages 1–6, 2008.

[140] C. Zhang, S. Pujades, M. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *Conference on Computer Vision and Pattern Recognition*, 2017.

[141] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. van Kaick, and A. Tagliasacchi. Deformation-driven shape correspondence. *Computer Graphics Forum*, 27(5):1431–1439, 2008.

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

[143] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Li. Face alignment across large poses: A 3d solution. In *Conference on Computer Vision and Pattern Recognition*, 2016.

[144] S. Zuffi and M. Black. The stitched puppet: A graphical model of 3d human shape and pose. In *Conference on Computer Vision and Pattern Recognition*, 2015.

A Publication list

International journals

- L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, B. Schiele. Building Statistical Shape Spaces for 3D Human Modeling. *Pattern Recognition*, 67:276–286, 2017.
- V. Fernández Abrevaya, S. Manandhar, F. Hétroy-Wheeler, S. Wuhrer. A 3D+t Laplace operator for temporal mesh sequences. *Computers & Graphics, Special Issue SMI'16*, 58:12–22, 2016.
- P. Kamousi, S. Lazard, A. Maheshwari, S. Wuhrer. Analysis of Farthest Point Sampling for Approximating Geodesics in a Graph. *Computational Geometry - Theory and Applications*, 57:1–7, 2016.
- T. Bolkart, S. Wuhrer. 3D Faces in Motion: Fully Automatic Registration and Statistical Analysis. *Computer Vision and Image Understanding*, 131:100–115, 2015.
- S. Wuhrer, J. Lang, M. Tekieh, C. Shu. Finite Element Based Tracking of Deforming Surfaces. *Graphical Models*, 77(1):1-17, 2015.
- A. Brunton, A. Salazar, T. Bolkart, S. Wuhrer. Review of Statistical Shape Spaces for 3D Data with Comparative Analysis for Human Faces. *Computer Vision and Image Understanding*, 128:1–17, 2014.
- S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, J. Lang. Estimation of Human Body Shape and Posture Under Clothing. *Computer Vision and Image Understanding*, 127:31–42, 2014.
- A. Salazar, S. Wuhrer, C. Shu, F. Prieto. Fully Automatic Expression-Invariant Face Correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014.
- A. Brunton, M. Wand, S. Wuhrer, H.-P. Seidel, T. Weinkauf. A Low-Dimensional Representation for Robust Partial Isometric Correspondences Computation. *Graphical Models*, 76(2):70–85, 2014.

- S. Wuhrer, C. Shu. Estimating 3D Human Shapes From Measurements. *Machine Vision and Applications*, 24(6):1133–1147, 2013.
- G. Aloupis, N. Benbernou, M. Damian, E. D. Demaine, R. Flatland, J. Iacono, S. Wuhrer. Efficient Reconfiguration of Lattice-Based Modular Robots. *Computational Geometry - Theory and Applications*, 46(8):917–928, 2013.
- J. Boisvert, C. Shu, S. Wuhrer, P. Xi. Three-Dimensional Human Shape Inference from Silhouettes: Reconstruction and Validation. *Machine Vision and Applications*, 24(1):145–157, 2013.
- G. Aloupis, M. Damian, R. Flatland, M. Korman, Ö. Özkan, D. Rappaport, and S. Wuhrer. Establishing Strong Connectivity using Optimal Radius Half-Disk Antennas. *Computational Geometry - Theory and Applications*, 46(3):328–339, 2013.
- C. Shu, S. Wuhrer, P. Xi. 3D Anthropometric Data Processing. *International Journal of Human Factors Modelling and Simulation*, 3(2):133–146, 2012.
- S. Wuhrer, C. Shu, P. Bose. Automatically Creating Design Models from 3D Anthropometry Data. *Journal of Computing and Information Science in Engineering*, 12(4):041007, 2012.
- P. Bose, M. Damian, K. Douïeb, J. O'Rourke, B. Seamone, M. Smid, S. Wuhrer. Pi/2-Angle Yao Graphs are Spanners. *International Journal of Computational Geometry and Applications, Special Issue ISAAC'10*, 22(1):61–82, 2012.
- S. Wuhrer, P. Xi, C. Shu. Human Shape Correspondence with Automatically Predicted Landmarks. *Machine Vision and Applications*, 23(4):821–830, 2012.
- S. Wuhrer, C. Shu, P. Xi. Posture-Invariant Statistical Shape Analysis Using Laplace Operator. *Computers & Graphics, Special Issue SMI'12*, 36(5):410–416, 2012.
- S. Wuhrer, C. Shu, P. Xi. Landmark-Free Posture Invariant Human Shape Correspondence. *The Visual Computer*, 27(9):843-852, 2011.
- P. Bose, A. Maheshwari, C. Shu, S. Wuhrer. A Survey of Geodesic Paths on 3D Surfaces. *Computational Geometry - Theory and Applications*, 44(9):486–498, 2011.
- G. Aloupis, S. Collette, M. Damian, E. D. Demaine, D. El-Khechen, R. Flatland, S. Langerman, J. O'Rourke, V. Pinciu, S. Ramaswami, V. Sacristan, S. Wuhrer. Efficient Constant-Velocity Reconfiguration of Crystalline Robots. *Robotica*, 29(1):59–71, 2011.

- A. Brunton, S. Wuhrer, C. Shu, P. Bose, E. Demaine. Filling Holes in Triangular Meshes Using Digital Images by Curve Unfolding. *International Journal of Shape Modeling*, 16(1-2):151-171, 2010.
- S. Wuhrer, P. Bose, C. Shu, J. O'Rourke, A. Brunton. Morphing of Triangular Meshes in Shape Space. *International Journal of Shape Modeling*, 16(1-2):195–212, 2010.
- S. Wuhrer, A. Brunton. Segmenting Animated Objects Into Near-Rigid Components. *The Visual Computer*, 26(2):147–155, 2010.
- G. Aloupis, S. Collette, M. Damian, E. D. Demaine, R. Flatland, S. Langerman, J. O'Rourke, S. Ramaswami, V. Sacristan, S. Wuhrer. Linear reconfiguration of cube-style modular robots. *Computational Geometry - Theory and Applications*, 42(6,7):652–663, 2009.
- P. Bose, P. Morin, and M. Smid, S. Wuhrer. Rotationally Monotone Polygons. *Computational Geometry - Theory and Applications*, 42(5): 471–483, 2009.
- T. Asano, P. Bose, P. Carmi, A. Maheshwari, C. Shu, M. Smid, S. Wuhrer. A Linear-Space Algorithm for Distance Preserving Graph Embedding. *Computational Geometry - Theory and Applications*, 42(4):289–304, 2009.
- P. Bose, P. Morin, M. Smid, and S. Wuhrer. Clamshell Casting. *Algorithmica*, 55(4):666–702, 2009.
- R. Atanassov, P. Bose, M. Couture, A. Maheshwari, P. Morin, M. Paquette, M. Smid, S. Wuhrer. Algorithms for optimal outlier removal. *Journal of Discrete Algorithms*, 7(2):239–248, 2009.
- P. Bose, V. Dujmović, D. Krizanc, S. Langerman, P. Morin, D. Wood and S. Wuhrer. A Characterization of the Degree Sequence of 2-trees. *The Journal of Graph Theory*, 58(3):191–209, 2008.
- S. Wuhrer, C. Shu, Z. Ben Azouz, P. Bose. Posture Invariant Correspondence of Incomplete Triangular Manifolds. *International Journal of Shape Modeling*, 13(2):139–157, 2007.

Book chapters

- A. Brunton, A. Salazar, T. Bolkart, S. Wuhrer. Statistical Shape Spaces for 3D Data: A Review In Handbook of Pattern Recognition and Computer Vision 5th Edition by Chi Hau Chen (Editor), 2016.

- S. Wuhrer, L. German, B. Rosenhahn. Statistical Human Body Modeling. In Digital Representation of the Real World: How to Capture, Model, and Render Visual Reality by M. Magnor, C. Theobalt, O. Sorkine-Hornung, O. Grau (Editors), 2015.

International peer-reviewed conferences and workshops

- V. Fernández Abrevaya, S. Wuhrer, E. Boyer. Multilinear Autoencoder for 3D Face Model Learning. Accepted to *IEEE Winter Conference on Applications of Computer Vision*, 2018.
- A. Bas, W. Smith, T. Bolkart, S. Wuhrer. Fitting a 3D Morphable Model to Edges: A Comparison Between Hard and Soft Correspondences. In *ACCV Workshop on Workshop on Facial Informatics*, 2016.
- A. Shehu, J. Yang, J.-S. Franco, F. Hétroy-Wheeler, S. Wuhrer. Computing temporal alignments of human motion sequences in wide clothing using geodesic patches. In *3D Vision*, 2016.
- J. Yang, J.-S. Franco, F. Hétroy-Wheeler, S. Wuhrer. Estimation of Human Body Shape in Motion with Wide Clothing. In *European Conference on Computer Vision*, 2016.
- T. Bolkart, S. Wuhrer. A Robust Multilinear Model Learning Framework for 3D Faces. In *Conference on Computer Vision and Pattern Recognition*, 2016.
- T. Bolkart, S. Wuhrer. A Groupwise Multilinear Correspondence Optimization for 3D Faces. In *International Conference on Computer Vision*, 2015.
- T. Bolkart, P. Bose, C. Shu, S. Wuhrer. A General Framework to Generate Sizing Systems from 3D Motion Data Applied to Face Mask Design. In *3D Vision*, 2014.
- A. Shehu, A. Brunton, S. Wuhrer, M. Wand. Characterization of Partial Intrinsic Symmetries.In *ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, 2014.
- A. Brunton, T. Bolkart, S. Wuhrer. Multilinear Wavelets: A Statistical Shape Space for Human Faces. In *European Conference on Computer Vision*, 2014.
- A. Hewer, I. Steiner, S. Wuhrer. A hybrid approach to 3D tongue modeling from vocal tract MRI using unsupervised image segmentation and mesh deformation. In *Interspeech*, 2014.

- G. Aloupis, J. Iacono, S. Langerman, Ö. Özkan, S. Wuhrer. The Complexity of Order Type Isomorphism. In *Symposium on Discrete Algorithms*, 2014.
- T. Bolkart, S. Wuhrer. Statistical Analysis of 3D Faces in Motion. In *3D Vision*, 2013.
- S. Wuhrer, J. Lang, C. Shu. Tracking Complete Deformable Objects with Finite Elements. In *Conference on 3D Imaging Modeling Processing Visualization and Transmission*, 2012.
- Y. Yang, D. Günther, S. Wuhrer, A. Brunton, I. Ivrissimtzis, H.-P. Seidel, T. Weinkauf. Correspondences of Persistent Feature Points on Near-Isometric Surfaces. In *ECCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, volume 7583 of LNCS, 2012.
- P. Bose, M. Damian, K. Douïeb, J. O'Rourke, B. Seamone, M. Smid, S. Wuhrer. Pi/2-Angle Yao Graphs are Spanners. In *International Symposium on Algorithms and Computation*, volume 6507 of LNCS, 2010.
- S. Wuhrer, Z. Ben Azouz, C. Shu. Posture Invariant Surface Description and Feature Extraction. In *Conference on Computer Vision and Pattern Recognition*, 2010.
- S. Wuhrer, C. Shu, J. Boisvert, G. Godin, P. Xi. Bending Invariant Meshes and Application to Groupwise Correspondences. In *ICCV Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, 2009.
- S. Wuhrer, C. Shu, P. Bose. Posture Invariant Correspondence of Triangular Meshes in Shape Space. In *ICCV Workshop on 3-D Digital Imaging and Modeling*, 2009.
- G. Aloupis, N. Benbernou, M. Damian, E. Demaine, R. Flatland, J. Iacono, S. Wuhrer. Efficient Reconfiguration for Lattice-Based Modular Robots. In *European Conference on Mobile Robots*, 2009.
- S. Wuhrer, C. Shu, M. Rioux. Posture Invariant Gender Classification for 3D Human Models. In *CVPR Workshop on Biometrics*, 2009.
- A. Brunton, S. Wuhrer, C. Shu, P. Bose, E. Demaine. Filling Holes in Triangular Meshes by Curve Unfolding. In *International Conference on Shape Modeling and Applications*, 2009.
- G. Aloupis, S. Collette, M. Damian, E. Demaine, D. El-Khechen, R. Flatland, S. Langerman, J. O'Rourke, V. Pinciu, S. Ramaswami, V. Sacristan, S. Wuhrer. Realistic Reconfiguration of Crystalline (and Telecube) Robots. In *International Workshop on the Algorithmic Foundations of Robotics*, 2008.

- G. Aloupis, S. Collette, E. D. Demaine, S. Langerman, V. Sacristan, S. Wuhrer. Linear reconfiguration of cube-style modular robots using O(log n) parallel moves. In *International Symposium on Algorithms and Computation*, volume 5369 of LNCS, 2008.
- G. Aloupis, S. Collette, M. Damian, E. D. Demaine, R. Flatland, S. Langerman, J. O'Rourke, S. Ramaswami, V. Sacristan, S. Wuhrer. Linear reconfiguration of cube-style modular robots. In *International Symposium on Algorithms and Computation*, volume 4835 of LNCS, 2007.
- A. Brunton, S. Wuhrer, and C. Shu. Image-Based Model Completion. In *International Conference on 3-D Digital Imaging and Modeling*, 2007.
- P. Bose, V. Dujmović, D. Krizanc, S. Langerman, P. Morin, D. Wood and S. Wuhrer. A Characterization of the Degree Sequence of 2-trees. *Workshop on Analytic Algorithms & Combinatorics*, 2007.

National conferences and workshops

- S. Wuhrer, C. Shu. Shape From Suggestive Contours Using 3D Priors. In *Canadian Conference on Computer and Robot Vision*, 2012.
- G. Aloupis, M. Damian, R. Flatland, M. Korman, Ö. Özkan, D. Rappaport, S. Wuhrer. Establishing Strong Connectivity using Optimal Radius Half-Disk Antennas. In *Canadian Conference on Computational Geometry*, 2011.
- S. Wuhrer, Z. Ben Azouz, C. Shu. Semi-Automatic Prediction of Landmarks on Human Models in Varying Poses. In *Canadian Conference on Computer and Robot Vision*, 2010.
- P. Bose, J. O'Rourke, C. Shu, S. Wuhrer. Isometric Morphing of Triangular Meshes. In *Canadian Conference on Computational Geometry*, 2008.
- Z. Ben Azouz, P. Bose, C. Shu, and S. Wuhrer. Approximations of Geodesic Distances for Incomplete Triangular Manifolds. In *Canadian Conference on Computational Geometry*, 2007.
- T. Asano, P. Bose, P. Carmi, A. Maheshwari, C. Shu, M. Smid, and S. Wuhrer. Linear-Space Algorithms for Distance Preserving Embedding. In *Canadian Conference on Computational Geometry*, 2007.
- P. Bose, P. Morin, M. Smid, S. Wuhrer. Rotationally Monotone Polygons. In *Canadian Conference on Computational Geometry*, 2006.
- R. Atanassov, P. Morin, S. Wuhrer. Removing Outliers to Minimize Area and Perimeter. In *Canadian Conference on Computational Geometry*, 2006.