UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2021-072

Faculty of Life Sciences, Technology and Communication

DISSERTATION

Presented on 4 October 2021 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE

by

ALEXANDRE SAINT

born on 21/08/1988 in Brussels (Belgium)

AUTOMATIC ANALYSIS, REPRESENTATION AND RECONSTRUCTION OF TEXTURED 3D HUMAN SCANS

Dissertation Defence Jury

Dr Djamila Aouada, Supervisor
Professor, University of Luxembourg

Dr Björn OTTERSTEN, Chair
Professor, University of Luxembourg

Dr Miguel Angel OLIVARES MENDEZ, Vice-Chair
Professor, University of Luxembourg

Dr Stefano BERRETTI, Member
Professor, University of Florence

Dr David FOFI, Member
Professor, University of Burgundy

Abstract

Various practical applications in computer vision are related to the human body. These involve representing and modelling the body shape, pose, clothing and appearance with mathematical and statistical tools requiring datasets of examples, representative of the variation in the data. Three-dimensional (3D) data is especially important as it allows to simulate the physical world directly, for example to analyse and lift ambiguities in other prevalent data modalities, such as images. However, existing datasets of 3D human scans show limitations in their size, diversity, quality or annotation. This reduces their applicability in tackling research questions around the 3D human body. Two particular applications of interest that remain unanswered are the estimation of body shape under clothing, and the completion of textured shape of missing or defective data.

This thesis proposes three main contributions. First, 3DBodyTex, a dataset of 3D human scans, which complements alternative datasets with real scans, body and clothing scans, hundreds of subjects, high-resolution texture information, dense annotations and aligned body shapes under the clothing. The aim is to enable and facilitate new research possibilities with learning-based methods, in 3D or using derived modalities. Second, to build this dataset automatically from raw scans, multiple robust 3D processing methods are proposed. These involve pose estimation, pose fitting, tight body shape fitting, and body shape estimation under clothing. The proposed methods show competitive or improved results on existing benchmarks and new proposed benchmarks based on 3DBodyTex. In particular, an alternative method is proposed to estimate the body shape under clothing from a single scan. On independent benchmarks, it is competitive with, or better than, methods requiring a full time sequence of scans. Third, the task of shape and texture completion of 3D human scans is tackled. A new method is proposed that completes the shape and the texture sequentially, and automatically identifies the missing regions. In particular, partial convolutions are extended to texture images (UV maps) for inpainting the colour of a 3D scan using a convolutional neural network. A new benchmark, based on 3DBodyTex, is proposed for the evaluation.

Acknowledgements

I would like to thank my advisor, Prof. Djamila Aouada, for her guidance and patience, and Prof. Björn Ottersten, for his support and trust. I am also grateful to Prof. Stefano Berretti, for his availability and advice, and Gleb Gusev and Kseniya Cherenkova, for their insights, support and optimism. This thesis was made possible thanks to the funding of Artec 3D and the Luxembourg National Research Fund (FNR) (11806282).

This work would not have been possible without the hundreds of people who volunteered to get scanned, nor without my colleagues at SnT, who helped operate the scanners. I would like to offer my special thanks to Ilse Van Renterghem for leading the extra scanning at Artec. To my colleagues and friends, you made this experience enjoyable, thank you. I feel privileged to have worked with you, Anis Kacem, and I appreciate the assistance you provided.

To my closest ones, thank you for your constant support and encouragements.

Contents

List of Figures **i**

List of Tables **i**

List of Acronyms **v**

1 Introduction **1**

1.1 Motivation 1

1.2 Scope of the thesis 3

1.3 Objectives and contributions 4

1.3.1 3DBodyTex: A diversified dataset of textured 3D human scans 5

1.3.2 Automatic methods for annotating and registering 3D human scans . 5

1.3.3 Human 3D shape and texture completion 7

1.4 Publications 8

1.5 Outline 10

2 Background on statistical body shape models **12**

2.1 SCAPE body model 14

2.1.1 Landmark regressor 16

2.2 SMPL body model 16

2.3 3D-CODED body model 17

3 3DBodyTex dataset **19**

3.1 Related datasets of 3D human scans 19

3.2 Proposed dataset 23

3.2.1 Contents 23

3.2.2 Acquisition process 23

3.2.3 Automatic processing 27

3.3 Conclusion 29

4 Automatic pose estimation **32**

4.1 Introduction 32

4.2 Related work 34

4.3 Proposed automatic pose estimation and fitting 35

4.3.1 Landmark estimation 35

4.3.2 Proposed pose fitting with SCAPE 36

4.3.3 Proposed pose fitting with SMPL 39

4.4 Evaluation 40

4.4.1 Evaluation of the pose fitting 40

4.5 Conclusion 41

5 Robust body shape fitting **43**

5.1 Introduction 43

5.2 Proposed shape fitting with SCAPE 44

5.2.1 Formulation of the fitting problem 44

5.2.2 Solving the fitting problem 46

5.2.3 Vertex correspondence 46

5.3 Proposed shape fitting with SMPL 46

5.4 Evaluation of the shape fitting with SCAPE 47

5.4.1 Qualitative evaluation 47

5.4.2 Quantitative evaluation 48

5.5 Evaluation of the shape fitting with SMPL 52

5.6 Conclusion 54

6 Body shape estimation under clothing **55**

6.1 Introduction 55

6.2 Related work 57

6.3 Proposed body shape estimation under clothing 58

6.3.1 Automatic segmentation 58

6.3.2 Estimating the body shape inside the clothing 60

6.3.3 Fitting multiple scans jointly 61

6.4 Evaluation 62

6.4.1 Evaluation of the automatic segmentation 62

6.4.2 Evaluation of the body shape estimation under clothing 63

6.5 Conclusion 65

7 Completion of partial textured 3D human scans **68**

7.1 Introduction 68

7.2 Related work 69

7.2.1 Shape completion 69

7.2.2 Texture completion 71

7.3 Problem formulation 72

7.3.1 Texture atlas 72

7.4 Proposed method: 3DBooSTeR 72

7.4.1 3D Body Shape Completion 74

7.4.2 Body Texture Completion 76

7.5 Proposed benchmark 80

7.5.1 Generating partial scans with ground truth 80

7.5.2 Evaluation metric for shape and texture 80

7.6 Evaluation 83

7.6.1 Shape completion 84

7.6.2 Texture transfer 85

7.6.3 Texture inpainting 86
7.7 Conclusion 87

8 Conclusion **89**

Glossary **91**

Bibliography **92**

List of Acronyms

3DBooSTeR *3D Body Shape and Texture Recovery* method [115]. 71

BPDN basis pursuit denoising [22] 38

IMU inertial measurement unit 22, 34

IRLS iteratively-reweighted least squares [54] 36

LBS linear blend skinning [84] 16, 17, 52, 53

SCAPE *Shape Completion and Animation of PoEple* parametric body shape model [5]. vi, xi, 6, 12, 13, 14, 15, 16, 17, 35, 40, 44, 50, 52, 53, 54

SHARP *SHApe Recovery from Partial textured 3D scans* Challenge and Workshop, ECCV 2020 [116]. x, 8, 28

SMPL *Skinned Multi-Person Linear* parametric body shape model [80]. xi, 6, 12, 13, 16, 17, 35, 40, 44, 50, 53, 54

VR virtual reality 1

List of Figures

2.1 Manually-defined skeleton for the Shape Completion and Animation of PoEple (SCAPE) body model (Section 2.1). The skeleton (black lines) is a tree structure following the main mostly-rigid body parts (coloured) separated by joints (white disks). The root node of the tree is at the joint on the upper part of the pelvis. 13

2.2 Triangle deformation process in the SCAPE body model (Section 2.1). A triangle (x_i^0, x_j^0, x_k^0) of the template mesh is deformed, in a local frame of reference around vertex x_i^0, by applying the learned transfromations R, Q, S (Equation (2.4)) to the two edges $\Delta x_{ij}^0, \Delta x_{ik}^0$ of the triangle (Equation (2.3)). 14

2.3 Triangle deformation process of the SCAPE body model [5]. The template is deformed into a pose, r, and shape, s, in two steps. First, each triangle is transformed individually (Equation (2.4) and Figure 2.2). Then, the triangles are reconnected by solving an optimisation problem (Equation (2.5)). (Description in Section 2.1.) 16

2.4 The encoder-decoder architecture of the 3D-CODED body model (Section 2.3). It is a deep learning model based on the PointNet [106] architecture and convolution operation, acting on unordered point sets. The decoder, f_d, can be seen as a body model transforming a template mesh, x^0, into a deformed mesh, x, using pose and shape parameters confounded into a single latent representation z. The encoder, f_e, allows to estimate the latent represenation of an input point set, y, even with partial data. This is exploited in Chapter 7 for shape completion. 18

3.1 Example scans of 3DBodyTex (row 1), with estimated segmentations (row 2), showing the variety in subjects, poses, and clothing. Section 3.2 describes the dataset, and Section 6.3.1, the segmentation method. 23

3.2 The Shapify Booth [6], a full-body 3D scanner by Artec 3D, is the system used to acquire the scans of 3DBodyTex (Section 3.2.2.1). 25

3.3 Samples of 3DBodyTex-2 in casual (top rows) and fitness (bottom row) clothing. For each person, ground-truth scan (left), sample synthetic partial scan (right). 27

3.4 Proposed pipeline for fitting a body model to a single scan presented in Chapters 4, 5 and 6. This serves as a base pipeline for creating the 3DBodyTex dataset (Section 3.2.3). Variations in the shape fitting allow to fit body scans tigtly (Chapter 5) or inside clothed scans (Chapter 6). The pose fitting (Chapter 4), the detection of landmarks (Section 4.3.1), and the segmentation (Section 6.3.1) are common to all types of scans. 30

5.1 Result of the fitting on a challenging pose. Left: Target scan. Middle: Target scan and fitted model overlaid. Right: Fitted model. 48

5.2 Fitting error for some sample scans of 3DBodyTex-1 in the U pose and some non-trivial poses. Left: Target scan. Middle and right: Fitting error as measured from scan to model for the front and back views, respectively. The colour scale varies from 0 cm to 1 cm. Values over 1 cm are coloured white. This is the case for regions discarded by the fitting because not part of the body shape, such as protruding hair or opened hands (since the body model has clenched fists). 49
5.3 Fitting error (mm) per body part on 3DBodyTex for BODYFIT and BODYFITR. Central bar: median. L: left, R: right. 51
5.4 Fitting error (mm) across all scans reported per vertex for BODYFIT (left) and BODYFITR (right) on the full body and with close-ups on the chest and right foot. 51
5.5 Example tight fitting results using BODYFITR (Sections 5.2 and 5.4) on challenging poses of 3DBodyTex. Back: scan. Front: fitted model. Middle: overlay. 52
5.6 Texture transfer from scan to fitted body model and examples of reposed model with texture (from left to right). 52
5.7 Fitting error per vertex aggregated across all scans for the proposed approach on 3DBodyTex. Left to right: median, mean, standard deviation. 53

6.1 Pipeline for estimating the segmentation of a 3D scan through 2D projections (Section 6.3.1). 59

7.1 Overview of the proposed approach for completing a partial textured 3D body mesh. 1) The complete shape is estimated. 2) The partial texture is transferred onto the estimated shape. 3) The corresponding texture image is inpainted. 69
7.2 Example of complete and partial texture atlases with their corresponding background masks M_b. Sample from the 3DBodyTex-2 dataset. 73
7.3 Flowchart of the proposed approach for 3D body shape and texture completion. 73

7.4 Pipeline for 3D shape completion: 1) An encoder-decoder network produces a first estimate of a complete shape. 2) The estimate is refined to better fit the clothing. 74
7.5 Identification of missing regions on a partial texture and binary masks calculation. 77
7.6 Results of shape completion for 6 examples of the test set. From left to right: input partial shape (white), initial shape estimate (orange), refined shape estimate (green), ground truth (white). In the input partial shapes, the visible interior surface is rendered in black. In the input partial shapes, the interior surface exposed in holes is rendered in black. 84
7.7 Shape refinement with symmetric versus directed (one-way) Chamfer distance for three examples of the test set. The one-way Chamfer distance (green) is the one retained in the proposed approach. From left to right: input partial shape (white), refined shape with symmetric Chamfer distance (orange), refined shape with directed (one-way) Chamfer distance (green), ground truth (white). In the input partial shapes, the visible interior surface is rendered in black. 85
7.8 Illustration of the texture transfer (right) from the partial shape (left) onto the refined shape estimate (orange, middle) for three examples of the test set. 85
7.9 Results for texture inpainting using different strategies. 86

List of Tables

3.1 Comparison of the main datasets of static 3D body scans. 20

3.2 Comparison of the main datasets of 3D clothed human scans. 21

3.3 Contents of the proposed 3DBodyTex dataset, described in Section 3.2.1. . . 24

3.4 Contents of the proposed 3DBodyTex dataset at three incremental stages of the construction and release process. The number of samples is reported per categorical and benchmark subsets, as used in the shape completion benchmark (Chapter 7) and the SHApe Recovery from Partial textured 3D scans (SHARP) Challenge [116]. The *standard poses* are the A and U rest poses. The *other poses* are varied, from a predefined list or arbitrary. The *casual* and *fitness* clothing types are shown in Figure 3.3. The per-scan *annotations* are landmarks, segmentations, registrations and texture maps (Table 3.3). . . . 28

4.1 Evaluation of the success rate of the proposed automatic pose fitting method (Sections 4.3.1 and 4.3) on over 2000 scans of 3DBodyTex (pairs of fitness-casual clothing), compared to the state of the art. (Discussion in Section 4.4.)
[a]: Applied only on the first 400 scans because of slow runtime, around 650 seconds per scan. 41

4.2 The main sources of failure, with observed frequency, of the proposed pose fitting method (Sections 4.3.1 and 4.3). The *highly-rotated limbs* are in the U and kneeling poses (Figure 3.1). (Discussion in Section 4.4.) 41

5.1 Root-mean-square vertex error (millimetres) of fitting tightly the scans in fitness clothing from 3DBodyTex, with the proposed approaches (Sections 5.2 and 5.3), incrementally improving. The methods using SCAPE fit the details non-rigidly and achieve lower errors, but also fit artefacts (Section 5.4). The method using Skinned Multi-Person Linear (SMPL) has higher errors, but avoids artefacts and is robust to irregularities in the shape, providing a globally-better fitting (Section 5.5). 50

6.1 Quantitative evaluation of the proposed automatic segmentation (Section 6.3.1) into coarse (highlighted) and fine categories. The metrics, *precision*, *recall* and *f-score* are Equations (6.9), (6.10) and (6.11). The *proportion* of each category corresponds to the weight of Equation (6.12). (Discussion in Section 6.4.1.) [a]: Does not appear in the data. 64

6.2 Root-mean-square vertex error (millimetres) of the proposed body shape estimation under the clothing on 3DBodyTex-3 (introduced in Section 6.3.2 and evaluated in Section 6.4.2). 64

6.3 Quantitative evaluation of the estimation of the body shape under the clothing on BUFF [156]. Root-mean-square error (millimetres) of the point-to-surface distance from the posed ground-truth mesh to the estimated mesh. For each motion, the best results from all other methods are highlighted, and also the best results of the proposed method. Different subjects indicated by ID number, *e.g.* 00005, one per column. (Reproduced from [156, Table 1, Section 6.2].) 66