Computer Vision, Imaging and Machine Intelligence Research Group (CVI^2)
Interdisciplinary Centre for Security, Reliability and Trust (SnT)
Faculty of Life Sciences, Technology and Communication (FSTM)
University of Luxembourg

Automatic Analysis, Representation and Reconstruction of Textured 3D Human Scans

PhD Thesis Defence

Alexandre Saint

04 October 2021

SnT, JFK Building – 29, Av. J.F. Kennedy, Luxembourg

Supervisor: Prof Dr Djamila Aouada

Outline

Outline

Humans in computer vision

Humans = centre of attention

- **applications**: assistance, detection, tracking, recognition, reconstruction, digital representation, animation, editing
- **domains**: autonomous driving, healthcare, medical rehabilitation, biomechanics, sports, security, entertainment, online retail
- **sensors**: RGB/infrared/depth cameras, laser scanners, lidar
- **modalities**: 2D (RGB, infrared), 2.5D (RGB-D), 3D (point cloud, mesh), 4D (3D + time)
- **features**: pose, motion, shape, appearance (colour/texture), interaction

Why 3D data?

Common modalities

	2D	2.5D	3D		4D
	RGB	RGB-D	point cloud	mesh	mesh sequence
memory efficiency	●	●	●	•	·
structure (grid)	●	•	·	·	·
3D interpretability	·	•	•	●	●
texture (colour)	●	•	·	•	•

better ● • · worse

Importance of 3D data

- Learn **accurate** models of the physical world.
- Serve as **prior** for lower-dimensional modalities (2D/2.5D).
- **Simulate** the physical world.
- **Generate** synthetic data for most modalities (2D to 4D).

But, 3D data is not perfect!

Defects

- shape artefacts
- texture artefacts
- missing data

Defects

- shape artefacts
- texture artefacts
- missing data

unresolved details
occlusion
material properties
texture interference

Defects

- shape artefacts
- texture artefacts
- missing data

unresolved details
occlusion
material properties
texture interference

Partial data.

Sources of partial 3D data

- **defects**
- partial acquisitions: to **accelerate** the acquisition time

Research questions?

Research questions

Body shape under clothing

Can we estimate an **accurate body shape under clothing** of a 3D scan?

1. From a **single scan**?
2. **No** exact **ground-truth** pairs body/clothing of real scans. How to alleviate?
3. How to exploit the **colour/texture**?
4. With **robustness** to **shape** variations: subjects, clothing?
5. With robustness to pose? **Challenging poses**?

Research questions

Shape and texture completion

Can we **complete** and correct the **shape and texture** of a **partial** 3D human **scan**?

1. Can one **identify** the regions to be completed? Robustly? Automatically?
2. Can one **recover** correct data by completion?
3. Does it allow a **reduction in scanning time** for an increase in computing resources?
4. **How much** missing **data** can be handled?

Outline

Introduction

Contributions

3DBodyTex dataset

Body shape estimation under clothing

Completion of shape and texture of partial 3D human scans

Conclusion

Research contributions

3DBodyTex **dataset**

1. **raw** scans: high-resolution, texture, real scans, variety
2. **generated** data: **aligned** body-clothing pairs, annotations
3. an automatic **pipeline** to generate the dataset from raw scans

Shape estimation **under clothing**

1. estimation of the **body shape under clothing** from a **single 3D scan**
2. **automatic** body model **fitting** of 3D human scans in **challenging poses**
3. method for **segmenting** 3D human scans with **minimal assumptions** on clothing types and colours

Shape and texture **completion** of 3D human scans

1. completion of partial textured mesh with **texture image inpainting**
2. **new evaluation metrics** for shape and texture completion

Outline

Why another dataset?

Motivation

Acquire **data for body shape estimation under clothing**

- approximately-aligned **ground-truth** pairs of body-clothing
- large **number of samples** for evaluation and supervision
- **variety** in subjects, poses, clothing
- high-quality **colour/texture**

Why *else* another dataset?

Motivation

Acquire **data for texture and shape completion**

- focus on **colour/texture**: variation and high-quality
- **variety** in subjects, poses, clothing
- different kinds of **defects** (scanning system: Shapify Booth[1])

[1] *Artec Shapify Booth* n.d.

What do existing datasets provide?

Related datasets

Body scans

criterion	**3DBodyTex-3**	SCAPE[2]	SPRING[3]	MPI[4]	MPII[5]	FAUST[6]	K3D-hub[7]	TOSCA[8]	CAESAR[9]
colour	texture	✗	✗	✗	✗	✗	✗	✗	vertex
colour resolution	high	-	-	-	-	-	-	-	low
with landmarks	✓	✗	✗	✗	✗	✗	✓	✗	✓
regular topology	✓	✗	✗	✓	✓	✗	✓	✓	✓
# poses	>35	70	1	35	1	30	5	20	3
# subjects	>500	1	≈3000	114	≈4300	10	50	3	4400
raw scans	✓	✗	✗	✓	✓	✗	✓	-	✓
# scans	3000	-	-	520	-	300	250	-	≈12000
resolution	high	-	-	high	-	high	low	-	high
watertight	✓	-	-	✗	-	✗	✗	-	✗
registered template	✓	✓	✓	✓	✓	✓(*)	✗	✓	✗
# registrations	3000	70	≈3000	520	4300	100	-	39	-
real people	✓	✓	✓	✓	✓	✓	✓	✗	✓
freely available	✓	✓	✓	✓	✓	✓	✓	✓	✗

(Saint, Ahmed, et al. 2018)
(Saint, Kacem, Cherenkova, and Aouada n.d.)

[2] Anguelov et al. 2005.

[3] Y. Yang et al. 2014.

[4] Hasler, Stoll, Sunkel, et al. 2009.

[5] Pishchulin et al. 2017.

Related datasets

Clothing scans

criterion	**3DBodyTex-3**	Vlasic *et al.*[10]	PerfCap[11]	Dressed Human[12]	BUFF[13]	CAPE[14] + ClothCap[15]	3DPW[16]	THuman[17]	SIZER[18]
data source	real	real	real	real	real	real[j]	real	real	real
ground-truth body	✓	✗	✗	✓	✓	✓	✓	✓[a]	✗
shape resolution	high	low	low	low	high	high	medium	low	high
sequence	✗	✓	✓	✓	✓	✓	✓	✗	✗
colour	texture	✗	✗	✗	vertex	vertex	✗	vertex	vertex
#subjects	>500	3		6	5	15	18	230	100
#clothes	>500	2		18	6	8	18	270	10
#poses	≤35	-	-	-	-	-	-	30[d]	1 ("A")
#motions	-	≤4		3	3	many	60	-	-
registered body	✓	-	-	✗	✗	✓[i]	✗	✓[a]	✓[c]
registered clothing	✗	✓[b]	✓[b]	✗	✗	✓	✓	✗	✓

(Saint, Kacem, Cherenkova, and Aouada n.d.)

[10] Vlasic et al. 2008.
[11] De Aguiar et al. 2008.
[12] J. Yang et al. 2016.
[13] C. Zhang et al. 2017.
[14] Ma et al. 2019.
[15] Pons-Moll et al. 2017.

What do we propose instead?

Proposed dataset

Proposed dataset

Proposed dataset

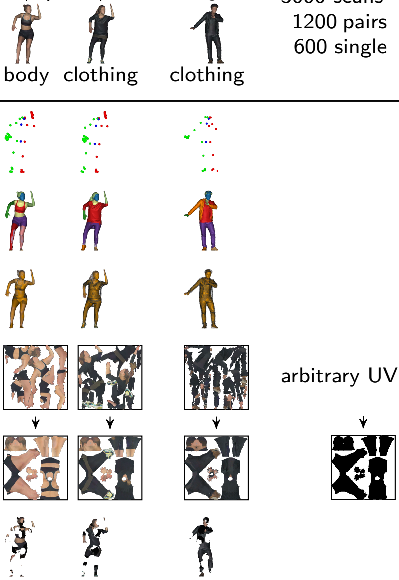
raw — scans

pair: body, clothing — single: clothing

3000 scans
1200 pairs
600 single

generated
- landmarks
- segmentation
- aligned body shapes
- transferred textures
- partial scans

arbitrary UV map



How do we create it?

Multiple acquisition phases

Proposed dataset
Acquisition

Phase 1

- clothing only
- 3 months

Phase 2.1

- pairs body-clothing
- about 1 year

Phase 2.2

- pairs body-clothing
- 3 weeks

Incremental creation

Incremental creation

3DBodyTex-1

Incremental creation

3DBodyTex-1

3DBodyTex-2

Incremental creation

3DBodyTex-1

3DBodyTex-2

3DBodyTex-3

Proposed dataset

Creation

3DBodyTex-1

- **body** scans
- 200 subjects

(Saint, Ahmed, et al. 2018)

Proposed dataset

Creation

3DBodyTex-2

- raw scans **body and clothing**
- 3000 scans
- **partial** scans (SHARP Challenge[21])

(Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020)

[21] Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020.

Proposed dataset

Creation

3DBodyTex-3

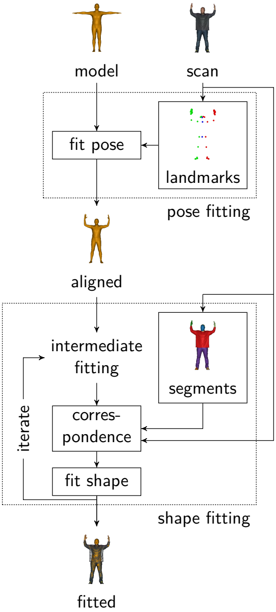
- **aligned pairs** body-clothing
- **1200** pairs = 2400 scans / 3000 scans
- **annotations**: segmentation, landmarks, registrations, common texture template

(Saint, Kacem, Cherenkova, and Aouada n.d.)

How was it processed?

Proposed dataset

Processing



Summary

3DBodyTex dataset brings

- **high resolution**: colour + shape
- 3000 scans
- **real** scans
- **variety**: subjects, poses, clothing
- 1200 raw **ground-truth pairs** body-clothing: approximately-aligned
- estimated **reference pairs** body-clothing: aligned
- **annotations**: sparse and dense

Summary

Applications

- **evaluation**
- **supervision**
- **simulation**
- **synthesis** of new data (from 2D to 4D)

Summary

Publications

- Saint, Ahmed, et al. 2018
- Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020
- Saint, Kacem, Cherenkova, and Aouada n.d.

Let's use the data!

Outline

Introduction

Contributions

3DBodyTex dataset

Body shape estimation under clothing

Completion of shape and texture of partial 3D human scans

Conclusion

Why estimate the body shape?

Motivation

Design **fitted clothing** from body measurements or shape.

Provide **non-invasive** scanning.

Monitor the body shape over time.

Physical ground truth: Animate, simulate clothed scans with correct physical constraints.

(online) retail

fitness

healthcare

entertainment

sports

ergonomics

safety

State of the art

body model-based fitting

(Hasler, Stoll, Rosenhahn, et al. 2009; Hirshberg et al. 2012; Bogo et al. 2014; Wuhrer et al. 2014; J. Yang et al. 2016; Pishchulin et al. 2017; C. Zhang et al. 2017)

+ regularising
- iterative
- lack degrees of freedom

learning-based approach

(Groueix et al. 2018; Deprelle et al. 2019; Hu et al. 2020)

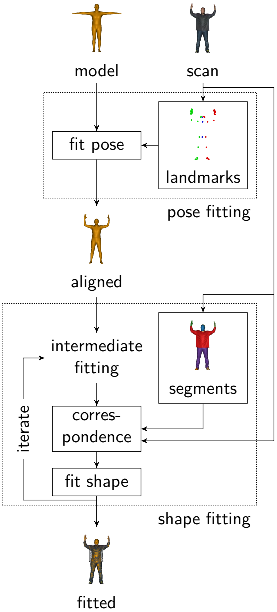
+ direct inference
- dependent on data (pose)
- lack accuracy (refinement)
- texture?

Objectives

- **accurate** shape estimation
- from **single scan**
- from **any scanning system**
- use **texture** if available

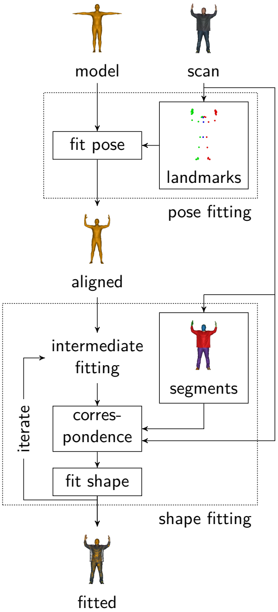
What do we propose?

Proposed fitting pipeline



Where do we start?

Pose estimation and fitting



Pose estimation and fitting

Why?

- Pose is (mostly) **independent** from the shape.
- But shape analysis **requires the pose**.
- Pose is key for **automating** the shape estimation under clothing.
  3DBodyTex has over **1000 independent** scans, in **uncontrolled** poses.

Pose estimation and fitting

State of the art

Hand-picked surface landmarks
(Wuhrer et al. 2014)

- manual
- clothing?

Motion capture (MoCap) surface landmarks
(Anguelov et al. 2005; J. Yang et al. 2016)

- specialised hardware

Predefined posture:
standing *A*-pose, front-facing
(Hirshberg et al. 2012; Pons-Moll et al. 2017; C. Zhang et al. 2017)

- limited poses

Global search
(Zuffi and Black 2015; J. Yang et al. 2016)

- slow (∼exhaustive search)
- success rate?

Regression of pose parameters from point cloud
(Groueix et al. 2018; Li et al. 2019)

- not using texture
- fails with extreme poses
- fails with wide clothing

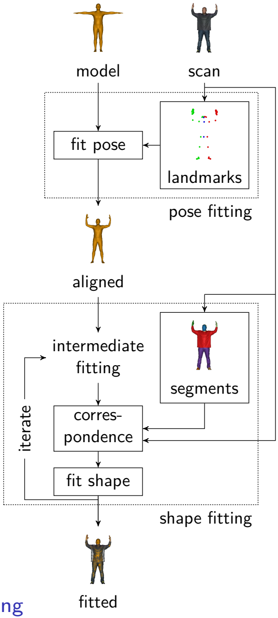
Pose estimation and fitting

Objectives

- from a single **3D scan** (with texture)
- **no additional information** (*e.g.* no MoCap landmarks)
- robust to **orientation**
- robust to **pose**
- robust to **clothing**
- robust to **subject**
- **automatic**

Pose estimation and fitting

Proposed approach

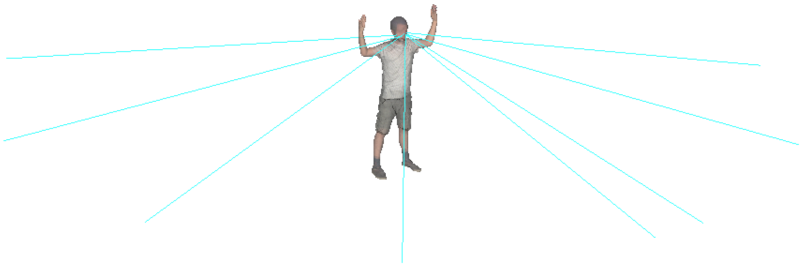


Pose estimation and fitting

Landmark estimation

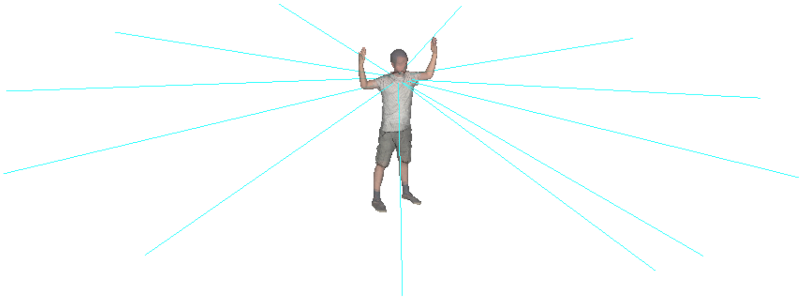
Pose estimation and fitting

Landmark estimation



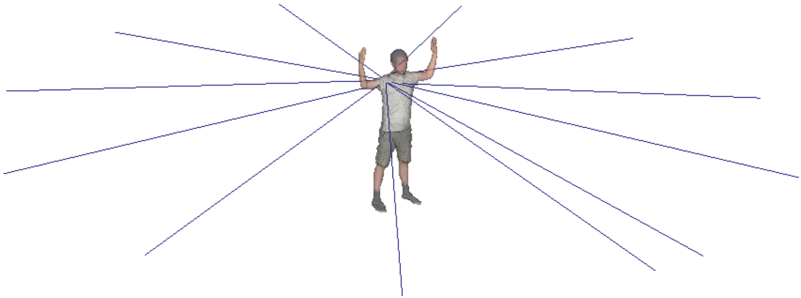
Pose estimation and fitting

Landmark estimation



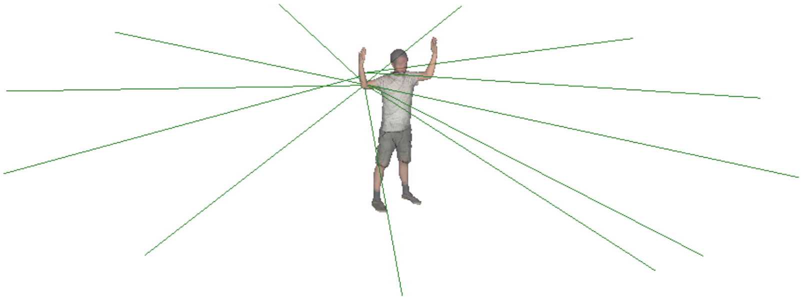
Pose estimation and fitting

Landmark estimation



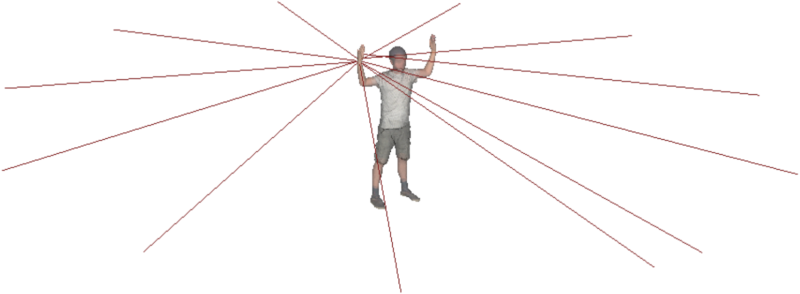
Pose estimation and fitting

Landmark estimation



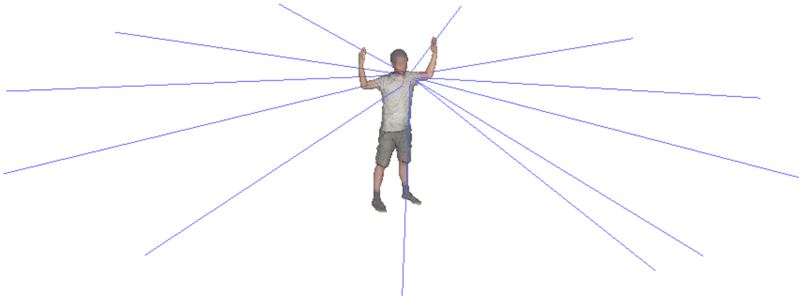
Pose estimation and fitting

Landmark estimation



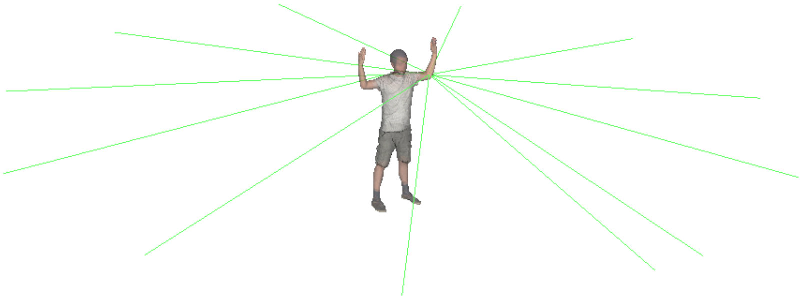
Pose estimation and fitting

Landmark estimation



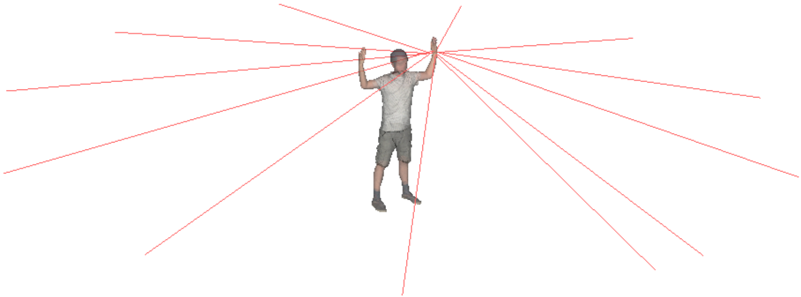
Pose estimation and fitting

Landmark estimation



Pose estimation and fitting

Landmark estimation

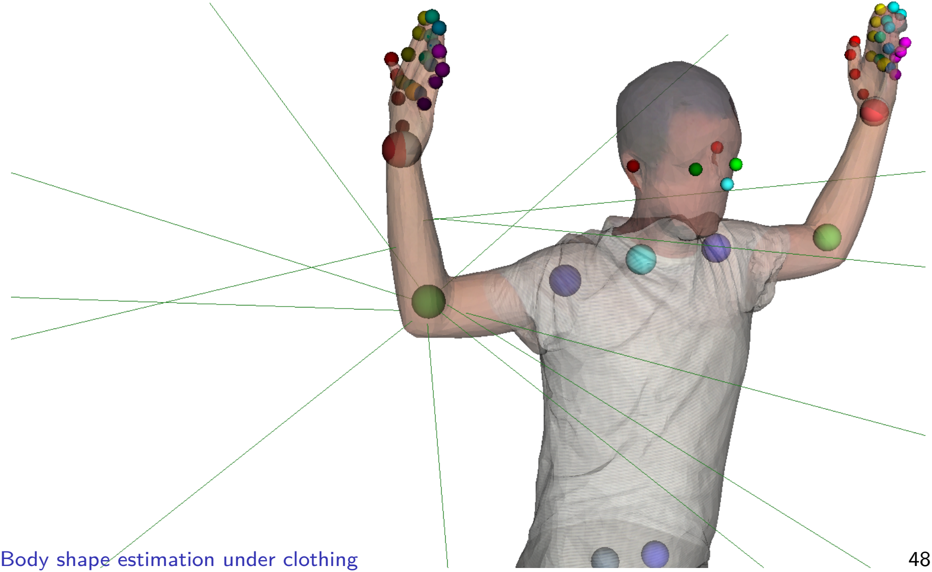


Pose estimation and fitting

Landmark estimation

Pose estimation and fitting

Landmark estimation



Pose estimation and fitting

Proposed approach

Estimate the **position** of a 3D landmark as the **best intersection** point, x, of multiple **rays**.

Data

- A single landmark, l.
- v views.
- One ray, r_v, per view.

Parametric equation of a ray

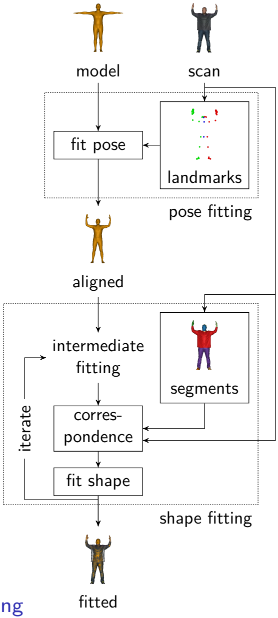
$$r(\alpha) = c + \alpha d \tag{1}$$

Optimisation problem

$$x^* = \arg\min_{x,\boldsymbol{\alpha}} \sum_{v \in \text{views}} \|x - (c^v + \alpha^v d^v)\|_1 \tag{2}$$

Pose estimation and fitting

Proposed approach



A **body model deforms** a **template body** mesh into a specific **pose** and **shape** (subject).

$M(r, s; x^0) = x(r, s; x^0)$	model
r	pose parameters
s	shape parameters
x^0	vertices of the template
x	deformed vertices
$M = \arg\min_x \sum_{(i,j)\in\text{edges}} \lVert RQS\Delta x^0_{ij} - \Delta x_{ij}\rVert_2^2$	SCAPE[21] **implicit**
$M = R\left(x^0 + S(s) + Q(r, s)\right)$	SMPL[22] **closed form**
R	pose transformations
S	shape transformations
Q	pose-dependent shape transfo.

[21] Anguelov et al. 2005.

[22] Loper et al. 2015.

Pose estimation and fitting

Proposed approach

Fit the landmarks by minimising

$$O_{\text{pose}}(r, s) = w_{\text{landmarks}} L_{\text{landmarks}} + w_{\text{pose}} P_{\text{pose}} + w_{\text{shape}} P_{\text{shape}} \tag{3}$$

where

$L_{\text{landmarks}}(r, s) = \frac{1}{L} \sum_l \|(\Lambda x - \hat{l})_l\|_2^2$	landmark fitting error
$\hat{l}$	estimated landmark position
Λ	landmark regressor
$x = x(r, s)$	deformed vertices of the model
$P_{\text{pose}}(r) = \frac{1}{J} \sum_j \|r_j\|_2^2$	pose prior
$P_{\text{shape}}(s) = \frac{1}{\lvert s \rvert} \sum_i s_i^2$	shape prior

How does it perform?

Success rate of pose fitting

Data

- about **2400 scans** from 3DBodyTex
- (pairs body-clothing)

Metric

- **success/failure**
- **visual** inspection

Results

method	success (%)
Zuffi and Black 2015[a]	27.3
proposed	**96.7**

[a]: on only 400 scans due to slow runtime (about 600s/scan)

(Saint, Kacem, Cherenkova, and Aouada n.d.)

Failure sources of pose fitting

failure source	failure rate (%)	
	relative	absolute
highly-flexed/rotated limb	**46.4**	**1.5**
landmark detection	**44.8**	**1.5**
scan defect	16.4	0.5
perpendicular hand	28.8	1.0
did not converge	8.5	0.3
all	100.0	**3.3**

Pose estimation and fitting

Summary

- use **texture** via 2D methods
- recovers **challenging poses**
- robust to **pose**, **clothing**, **subject**
- **automatic**
- **predictable/systematic**
  no randomness as in (Zuffi and Black 2015)

Publications

- Saint, Shabayek, et al. 2017
- Saint, Ahmed, et al. 2018
- Saint, Cherenkova, et al. 2019
- Saint, Kacem, Cherenkova, and Aouada n.d.

Pose estimation and fitting

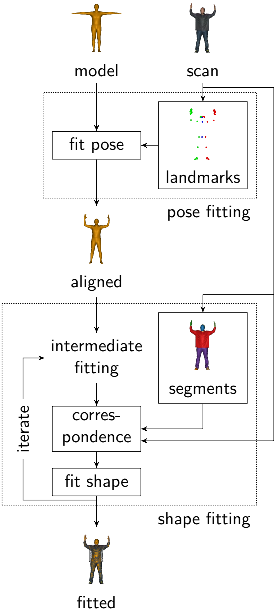
Limitations

- Less robust to **orientation of the hands**.
  Mitigated with other view angles, but then **challenging viewpoints**[23].
- Not robust to **partial** data: artefacts or missing data.
  (See *shape and texture completion* contribution.)

[23] Baptista et al. 2021.

What about the shape?

Tight body shape fitting



Why fit the body shape tightly?

Tight body shape fitting

Motivation

Our goals

- Obtain **ground truth** for evaluation.
- **Regularise/constrain** the shape estimation under clothing.

What we also get

- Dense surface **correspondence** (template registration): for learning from data (*e.g.* body model).
- Dense **annotations**: anatomical regions. (To generate the ground-truth data in the SHARP Challenge[24].)

[24] Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020.

Tight body shape fitting

State of the art

model-free template registration
(Allen, Curless, and Popović 2003; Hirshberg et al. 2012)

- heuristics to regularise

body model-based fitting
(Hasler, Stoll, Rosenhahn, et al. 2009; Hirshberg et al. 2012; Bogo et al. 2014; Wuhrer et al. 2014; Pishchulin et al. 2017)

- point-to-point loss
- stall in artefacts (outliers)

estimation of point correspondence
(Marin et al. 2018)

- similar mesh topology

Tight body shape fitting

Objectives

- **robust** to artefacts
- target **any** mesh **topology**

Tight body shape fitting

Proposed approach

Minimise

$$O_{\text{tight}}(r,s) = w_{\text{tight}} L_{\text{tight}} + w_{\text{pose}} P_{\text{pose}} + w_{\text{shape}} P_{\text{shape}} \tag{4}$$

where

$L_{\text{tight}}(r,s) = \frac{1}{\lvert C \rvert} \sum_{(x,y)\in C} \left\lVert \boldsymbol{n_y^T(x-y)} \right\rVert_{\boldsymbol{1}}$	**vertex fitting** error
$\left\lVert n_y^T(x-y) \right\rVert_2^2$	standard point-to-plane metric[25]
$\left\lVert \boldsymbol{n_y^T(x-y)} \right\rVert_{\boldsymbol{1}}$	more **robust** point-to-plane metric
$x = x(r,s)$	deformed vertex of the model
y	vertex on the scan
C	set of vertex correspondence

[25] Chen and Medioni 1992.

Tight body shape fitting

Proposed approach

method	model	L_{tight} distance
BODYFIT	SCAPE	$\|x - y\|_2^2$
		(Saint, Ahmed, et al. 2018)
equivalent to		(Hasler, Stoll, Rosenhahn, et al. 2009; Hirshberg et al. 2012; Bogo et al. 2014; Wuhrer et al. 2014; Pishchulin et al. 2017)
BODYFITR	SCAPE	$\left\| n_y^T (x - y) \right\|_2^2$ (standard point-to-plane[26])
more robust		(Saint, Cherenkova, et al. 2019)
BODYFITRR	SMPL	$\left\| \boldsymbol{n}_{\boldsymbol{y}}^{\boldsymbol{T}} (\boldsymbol{x} - \boldsymbol{y}) \right\|_{\boldsymbol{1}}$ (**proposed** point-to-plane)
more robust		(Saint, Kacem, Cherenkova, and Aouada n.d.)

[26] Chen and Medioni 1992.

Tight body shape fitting

Proposed approach

Alternative robust loss

- $\rho_\alpha \left(\|x - y\|_2^2 \right)$
- example: Geman-McClure[27] $\rho_\alpha(x) = \frac{2(x/\alpha)^2}{((x/\alpha)^2+4)}$
- (C. Zhang et al. 2017)
- Pro: Parameter α for flexibility.
- Con: Need to choose an α.

Proposed

- $\left\| n_y^T (x - y) \right\|_1$
- (Saint, Kacem, Cherenkova, and Aouada n.d.)
- Pro: No parameter to tune.
- Con: Maybe less flexibility.

[27] Ganan and McClure 1985.

Do we actually fit well?

Tight body shape fitting

Evaluation

Root-mean-square vertex error (millimetres)					
method	model	3DBodyTex-1		3DBodyTex-3	
		mean	median	mean	median
BODYFIT[28]	SCAPE	3.1	1.5	-	-
BODYFITR[29]	SCAPE	2.4	1.3	-	-
BODYFITRR[30]	SMPL	5.4	3.3	5.3	3.2

vertex correspondence: closest point

[28] Saint, Ahmed, et al. 2018.

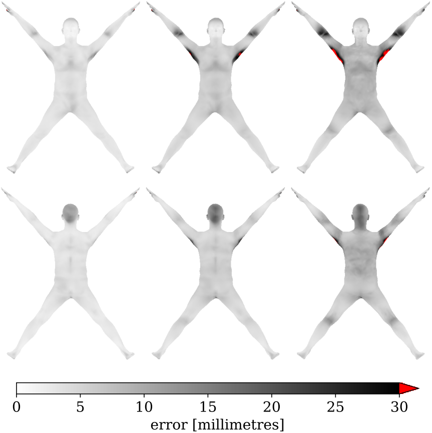
[29] Saint, Cherenkova, et al. 2019.

[30] Saint, Kacem, Cherenkova, and Aouada n.d.

Tight body shape fitting

Evaluation

median	mean	standard dev.
3.2	5.3	



Tight body shape fitting

Summary

- **comparison** of multiple **metrics** and **two body models**
- proposed an alternative **robust** loss
- **no parameter** to tune
- fit the **details** with SCAPE
- promote stronger **regularisation** of the model with SMPL

Publications

- Saint, Ahmed, et al. 2018
- Saint, Cherenkova, et al. 2019
- Saint, Kacem, Cherenkova, and Aouada n.d.

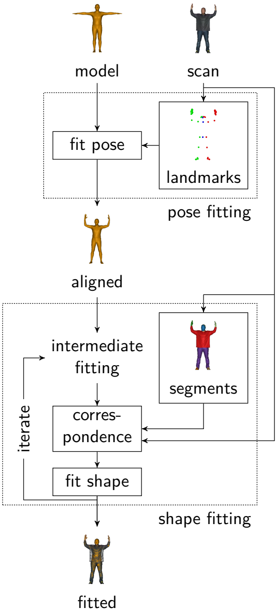
Tight body shape fitting

Limitations

- cannot tell the **difference** between **artefacts and details**

What if there is clothing?

Clothing and body segmentation



Why segment?

Clothing and body segmentation

Motivation

Segment to **adapt** the estimation of the body shape to **different regions**:

- clothing,
- visible skin,
- hair,
- artefacts.

State of the art

colour-based skin segmentation

(C. Zhang et al. 2017)

- only skin/no skin

2D clothing segmentation

(Gong et al. 2019)

2D-based 3D clothing segmentation

(Pons-Moll et al. 2017; Yu et al. 2018)

- uniform colours
- fixed clothing types

on point clouds

(Jertec et al. 2019)

- difficulty touching/fused parts (body, clothing)
- low resolution
- miss details

Clothing and body segmentation

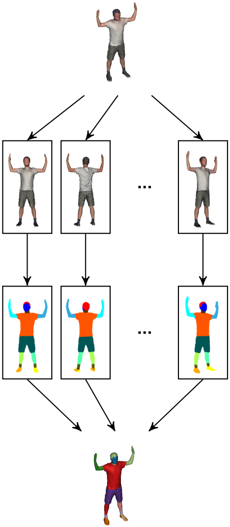
Objectives

Segment

- **any clothing** type
- **any colour** and patterns
- independently of **pose**
- with **touching/fused** shape

Clothing and body segmentation

Proposed approach



3D scan

2D views

2D segmentation
(Gong et al. 2019)

3D back-projection

How does it perform?

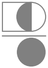
Clothing and body segmentation

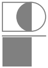
Evaluation

Data

- test ground truth: manual segmentation of 15 scans

Metrics

$P = w \frac{|S \cap \hat{S}|}{|\hat{S}|}$ precision (weighted) 

$R = w \frac{|S \cap \hat{S}|}{|S|}$ recall (weighted) 

$F = \frac{2PR}{P+R}$ f-score

S ground-truth segmentation 

$\hat{S}$ estimated segmentation 

$w_c = \frac{|S_c|}{|\cup_{c'} S_{c'}|}$ relative proportion of category c

Clothing and body segmentation
Evaluation

category coarse	precision	recall	**f-score**	proportion
skin	0.95	0.82	0.87	0.08
hair	0.85	0.89	0.87	0.07
clothing	0.77	0.84	0.81	0.84
all	0.79	0.84	**0.82**	1.00

Clothing and body segmentation

Summary

Proposed a method for **segmenting** a 3D human **scan** into **clothing** items and **body** regions:

- uses **texture** via 2D methods
- no assumption on **clothing types**
- no assumption on **clothing colours and texture**
- about **80%** accuracy
- **simple**

Publication

- Saint, Kacem, Cherenkova, and Aouada n.d.

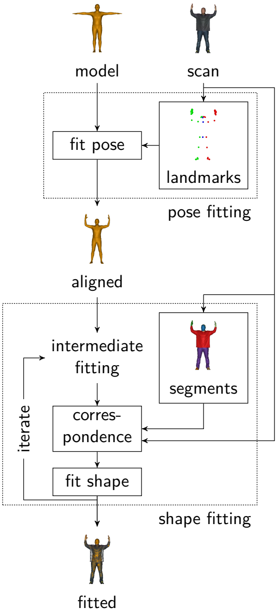
Clothing and body segmentation

Limitations

- relies on **texture**
- relies on the performance of the **underlying 2D method**
- **20% margin** for improvement
- simple **voting** procedure
- 2D and 3D stages **not integrated**
  improvement: learning-based end-to-end method combining 2D projections and class predictions

We have barely scraped the surface!

Shape estimation under clothing



Motivation

- **non-invasive** alternative to body scanning
- **fitted clothing** design
- **virtual** clothing **try-on**
- physical **simulation**: virtual avatar, ergonomics

State of the art

Single 3D scan

(Hasler, Stoll, Rosenhahn, et al. 2009; Wuhrer et al. 2014)

- qualitative evaluation
- quantitative evaluation on body measurements
- few subjects

Time sequence of scans (4D scans)

(Neophytou and Hilton 2014; Wuhrer et al. 2014; J. Yang et al. 2016; C. Zhang et al. 2017)

- require multiple scans

Real-time estimation (RGB-D stream)

(Yu et al. 2018)

- RGB-D
- focus real time
- not accurate

Direct inference (point cloud)

(Groueix et al. 2018; Hu et al. 2020)

- not accurate
- require refinement
- no texture

Objectives

- **accurate** shape estimation
- from **single scan**
- from **any scanning system**
- use **texture** if available

Proposed approach

Minimise

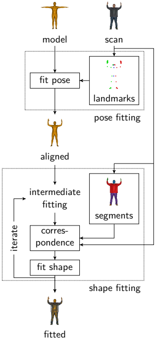
$$O_{\text{inside}}(r,s) = w_{\text{inside}} L_{\text{inside}} + w_{\text{tight}} L_{\text{tight}} + w_{\text{volume}} P_{\text{volume}} + w_{\text{pose}} P_{\text{pose}} + w_{\text{shape}} P_{\text{shape}} \tag{5}$$

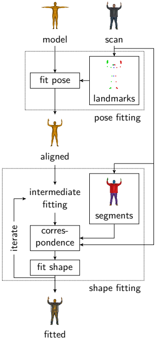
where

$L_{\text{inside}}(r,s) = \frac{1}{\sum_x o(x)} \sum_{(x,y)\in C} \boldsymbol{o}(\boldsymbol{x}) \left\| n_y^T (x-y) \right\|_1$	**inside fitting** error
$\boldsymbol{o}(\boldsymbol{x}) \in \{0,1\}$	inside $=0$, **outside** $=\mathbf{1}$
$\left\| n_y^T (x-y) \right\|_1$	**robust** point-to-plane metric
$L_{\text{tight}}(r,s) = \frac{1}{\lvert C \rvert} \sum_{(x,y)\in C} \left\| n_y^T (x-y) \right\|_1$	tight vertex fitting error
$P_{\text{volume}}(r,s) = \left(1 - \frac{V_{\text{model}}}{V_{\text{scan}}}\right)^2$	**volume** prior

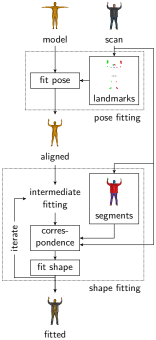
Let's reuse!

Shape estimation from multiple scans





. . .



Proposed approach

Joint fitting

Minimise

$$O_{\text{joint}} = \sum_i f_i O_{\text{tight},i} + \sum_i c_i O_{\text{inside},i} + w_{\text{subject}} \sum_{i<j} L_{\text{subject}}(s_i, s_j) \qquad (6)$$

where

$O_{\text{tight}}(r, s)$	fit body scans tightly
$\quad f_i \in \{0, 1\}$	not body scan $= 0$, body scan $= 1$
$O_{\text{inside}}(r, s)$	fit clothed scans inside
$\quad c_i \in \{0, 1\}$	not clothed scan $= 0$, clothed scan $= 1$
$L_{\text{subject}}(s, s') = \lVert s - s' \rVert_2^2$	enforce a single subject

Does this keep us covered?

Evaluation on 3DBodyTex

Root-mean-square vertex error (RSMVE) (millimetres)
vertex correspondence: closest point

method	3DBodyTex-3	
	mean	median
proposed	7.3	5.2

[31] Saint, Ahmed, et al. 2018.
[32] Saint, Cherenkova, et al. 2019.
[33] Saint, Kacem, Cherenkova, and Aouada n.d.

Evaluation on 3DBodyTex

Root-mean-square vertex error (RSMVE) (millimetres)
vertex correspondence: closest point

method	3DBodyTex-3	
	mean	median
proposed	7.3	5.2

Comparison with **tight body shape fitting**

method	model	3DBodyTex-1		3DBodyTex-3	
		mean	median	mean	median
BODYFIT[31]	SCAPE	3.1	1.5	-	-
BODYFITR[32]	SCAPE	2.4	1.3	-	-
BODYFITRR[33]	SMPL	5.4	3.3	5.3	3.2

[31] Saint, Ahmed, et al. 2018.
[32] Saint, Cherenkova, et al. 2019.
[33] Saint, Kacem, Cherenkova, and Aouada n.d.

Evaluation on BUFF

BUFF dataset (C. Zhang et al. 2017)

- time sequences of 3D scans (4D scans)
- 5 subjects
- 3 motions (about 10 seconds, 300 frames per motion)
- 2 clothing outfits
  casual long pants/t-shirt
  sports shorts/t-shirt

Related methods

- body shape estimationg under clothing from **4D scans**
- J. Yang et al. 2016
- C. Zhang et al. 2017: **detailed** shape

Evaluation on BUFF

Root-mean-square vertex error (millimetres)

method	variant	frames	t-shirt, long pants					soccer outfit				mean
			00005	00032	00096	00114	03223	00005	00032	00114	03223	
			motion: hips									
J. Yang et al. 2016		all	21.02	15.77	21.66	18.05	21.84	22.52	16.81	17.54	22.03	19.51
C. Zhang et al. 2017	fusion	all	2.81	2.66	2.71	2.65	2.54	2.65	2.63	2.57	2.50	2.63
	details	all	**2.75**	**2.63**	**2.64**	**2.56**	**2.40**	**2.58**	**2.59**	**2.51**	**2.38**	**2.55**
proposed	single	1	5.06	4.84	5.12	4.94	6.47	5.00	5.32	5.32	6.45	5.39
	joint	3	**3.21**	**3.50**	**4.67**	**3.34**	**4.43**	**3.17**	**3.58**	**3.32**	**4.52**	**3.75**
		6	3.41	3.58	5.10	3.55	4.53	3.23	3.61	3.46	4.69	3.90
		12	3.47	3.94	4.80	3.94	4.77	3.82	3.71	3.47	4.55	4.05

BUFF dataset (C. Zhang et al. 2017)

Evaluation on BUFF

Root-mean-square vertex error (millimetres)

method	variant	frames	t-shirt, long pants					soccer outfit				mean
			00005	00032	00096	00114	03223	00005	00032	00114	03223	
			motion: shoulders mill									
J. Yang et al. 2016		all	18.77	18.02	19.02	14.78	18.15	18.74	17.88	16.37	19.47	17.59
C. Zhang et al. 2017	fusion	all	2.56	2.74	2.92	2.69	2.42	2.89	2.87	2.58	2.44	2.63
	details	all	**2.49**	**2.72**	**2.85**	**2.59**	**2.26**	**2.83**	**2.82**	**2.51**	**2.33**	**2.55**
proposed	single	1	5.50	5.38	n/a[a]	5.17	5.63	5.49	5.78	5.29	5.33	5.45
	joint	3	**4.13**	**3.52**	n/a[a]	4.69	**4.29**	4.49	**3.83**	**3.64**	4.62	4.15
		6	4.18	3.53	n/a[a]	**3.76**	4.58	**3.50**	3.87	3.74	**4.47**	**3.96**
		12	**4.13**	3.60	n/a[a]	4.21	4.40	4.29	3.84	**3.64**	4.67	4.10

BUFF dataset (C. Zhang et al. 2017)

Evaluation on BUFF

Root-mean-square vertex error (millimetres)

method	variant	frames	t-shirt, long pants					soccer outfit				mean
			00005	00032	00096	00114	03223	00005	00032	00114	03223	
			motion: tilt twist left									
J. Yang et al. 2016		all	17.29	13.76	18.68	15.42	17.90	16.88	16.96	16.40	20.41	17.27
C. Zhang et al. 2017	fusion	all	2.58	2.39	2.89	2.38	2.43	2.50	2.63	2.28	2.28	2.47
	details	all	**2.52**	**2.36**	**2.83**	**2.31**	**2.27**	**2.44**	**2.59**	**2.23**	**2.17**	**2.40**
proposed	single	1	4.48	4.64	4.74	6.02	5.66	4.61	4.99	4.97	5.15	5.03
	joint	3	3.17	**3.63**	3.81	**3.87**	**4.41**	3.28	3.72	3.48	4.29	3.74
		6	**3.12**	3.64	**3.68**	4.06	**4.41**	3.20	**3.71**	**3.46**	4.20	**3.72**
		12	3.39	3.84	3.96	4.32	4.56	**3.18**	3.97	3.49	**4.18**	3.88

BUFF dataset (C. Zhang et al. 2017)

Summary

Proposed a method for shape estimation under clothing

- from a **single** scan
- **adaptive** fitting using segmented **texture**
- **comparable** to the state of the art on **time sequences**
- can fit **multiple** related scans jointly for increased performance
- **peak** performance with fitting **3 to 6 scans**, instead of a full sequence

Publication

- Saint, Kacem, Cherenkova, and Aouada n.d.

Limitations

- not fitting the shape **details**
- does not handle **partial** scans

...

Outline

Introduction

Contributions

3DBodyTex dataset

Body shape estimation under clothing

Completion of shape and texture of partial 3D human scans

Conclusion

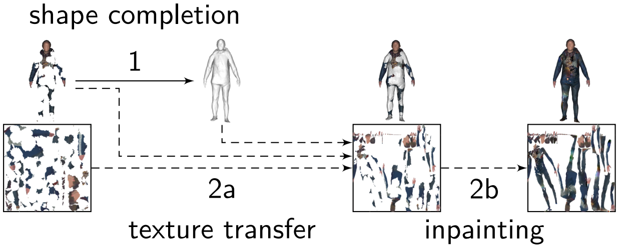
Motivation

- **correct** defects
- **compensate** for physical limitations of 3D scanners
- **reduce acquisition time** by scanning part of the target and completing the rest

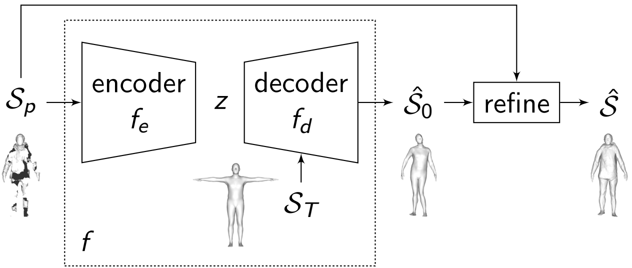
State of the art

hole-filling	+ simple
(Davis et al. 2002)	- small holes only
template/body model	+ regularisation
(Szeliski and Lavallée 1996; Anguelov et al. 2005)	- clothing?
volumetric convolutions	+ texture
(Wu et al. 2015; Dai, Ruizhongtai Qi, and Nießner 2017; Han et al. 2017)	- simple shapes - low resolution - high computational resources
point cloud convolutions	+ handle partial data
(Groueix et al. 2018; Deprelle et al. 2019)	- need fitting refinement
mesh convolutions	- need fitting refinement
(Litany et al. 2018; Ma et al. 2019)	- not robust to mesh topology
implicit-function CNN	+ texture
(Chibane, Alldieck, and Pons-Moll 2020)	- post-processing - medium resolution - high computational resources

Proposed approach

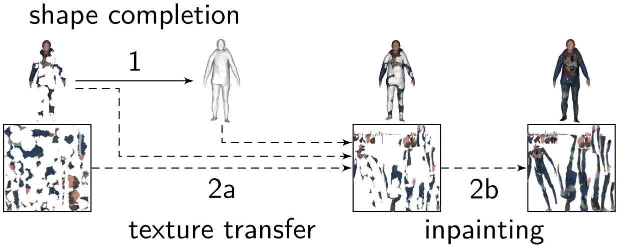


Shape completion

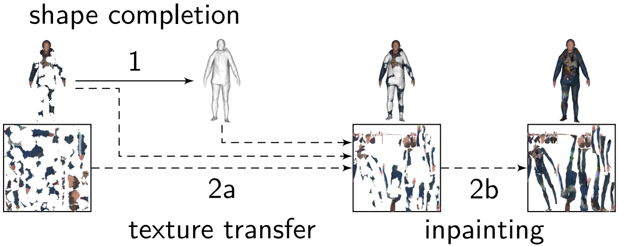


(Groueix et al. 2018; Deprelle et al. 2019)

Texture transfer



Texture completion



Identification of missing regions

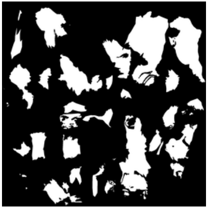
input


mesh


texture image


background mask
M'

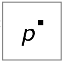
output


partial mask
M
missing regions

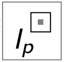
Inpainting

$y_p = w^T I_p$ — standard **convolution**

w — weights (convolution kernel)

$I =$ 

— image I, pixel $\boldsymbol{p}$

$I =$ 

— patch $\boldsymbol{I_p}$ around p

$y_p = w^T \left(\boldsymbol{M_p} \odot I_p \right) \frac{\boldsymbol{k^2}}{|\boldsymbol{M_p}|}$ — **partial** convolution

k — kernel size

$M =$  — partial mask

(Liu et al. 2018) Inpainting images with arbitrary holes

$y_p = w^T \left(M_p \odot \boldsymbol{M'_p} \odot I_p \right) \frac{k^2}{|M_p \odot \boldsymbol{M'_p}|}$ — partial convolution on **texture image**

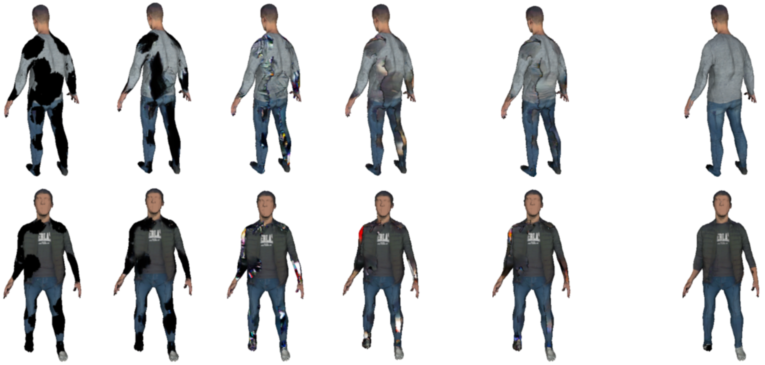
$M' =$  — background mask

3DBooSTeR (Saint, Kacem, Cherenkova, and Aouada 2020)

How does it perform?

Evaluation

Qualitative



input	IN	IN/M'	BT/M'	**IN+BT/M'**	ground truth

IN	training on ImageNet (Deng et al. 2009)
BT	training on 3DBodyTex
IN+BT	pretraining on ImageNet + fine-tuning on 3DBodyTex
/M'	with proposed texture image masked convolution

Evaluation

Quantitative

SHARP challenge and workshop

- (Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020)

Data

- Proposed 3DBodyTex-2.

Metrics

- Proposed specific metrics for evaluating shape and texture completion.

Summary

Method

- proposed a **sequential** method for shape and texture completion
- extended **masked convolutions** for inpainting **texture images** (Liu et al. 2018)

Publication

- Saint, Kacem, Cherenkova, and Aouada 2020

Evaluation

- proposed **evaluation metrics** for shape and texture completion for the **SHARP Challenge/Workshop**

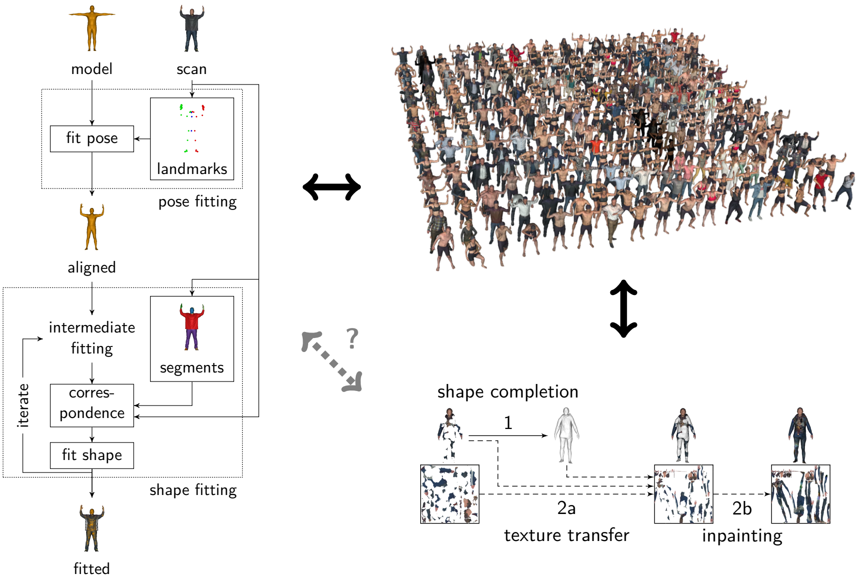
Publication

- Saint, Kacem, Cherenkova, Papadopoulos, et al. 2020

Completed!

Outline

Summary


model
scan
fit pose
landmarks
pose fitting
aligned
intermediate fitting
segments
iterate
corres-pondence
fit shape
shape fitting
fitted
?
shape completion
1
2a
2b
texture transfer
inpainting

Future directions

- Rely on **3DBodyTex** for learning? (weak-)supervision? generating ground-truth data?
- Can we **integrate** methods with sequential stages into methods **learnable** end to end?
  (pose estimation/fitting, segmentation, shape fitting/under clothing, completion)
- How to **learn** to estimate the body shape with related scans that are **not aligned**?

Publications I

Journals

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, and Djamila Aouada (n.d.). "3DBodyTex: Building a textured dataset of 3D human scans with aligned body shapes". In: (). To be submitted.

Conferences

Saint, Alexandre, Eman Ahmed, et al. (2018). "3dbodytex: Textured 3d body dataset". In: *2018 International Conference on 3D Vision (3DV)*. IEEE, pp. 495–504.

Saint, Alexandre, Kseniya Cherenkova, et al. (2019). "BODYFITR: Robust automatic 3D human body fitting". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 484–488.

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, and Djamila Aouada (2020). "3DBooSTeR: 3D Body Shape and Texture Recovery". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Publications II

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, Konstantinos Papadopoulos, et al. (2020). "SHARP 2020: The 1st Shape Recovery from Partial Textured 3D Scans Challenge Results". In: *Proceedings of the European Conference on Computer Vision (ECCV)*.

Saint, Alexandre, Abd El Rahman Shabayek, et al. (2017). "Towards automatic human body model fitting to a 3d scan". In: *Proceedings of 3DBODY. TECH 2017-8th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Montreal QC, Canada, 11-12 Oct. 2017*. Hometrica Consulting, pp. 274–280.

Publications not included in this thesis

Baptista, Renato et al. (2021). "Towards generalization of 3d human pose estimation in the wild". In: *International Conference on Pattern Recognition*. Springer, pp. 72–81.

Acknowledgements

This work was possible thanks to **a lot of contributors** and **supporters**:

- the **volunteers** for the scanning
- the **helpers** for the scanning (CVI^2 members, SnT members, externs)
- **Artec 3D**
- the **Luxembourg National Research Fund** (FNR)

Thank you!

References I

Allen, Brett, Brian Curless, and Zoran Popović (2003). "The space of human body shapes: reconstruction and parameterization from range scans". In: *ACM transactions on graphics (TOG)* 22.3, pp. 587–594.

Anguelov, Dragomir et al. (2005). "SCAPE: shape completion and animation of people". In: *ACM Transactions on Graphics (TOG)*. Vol. 24. 3. ACM, pp. 408–416.

Artec Shapify Booth (n.d.). `https://www.artec3d.com/portable-3d-scanners/shapifybooth`. Retrieved: 2021-06-06.

Baptista, Renato et al. (2021). "Towards generalization of 3d human pose estimation in the wild". In: *International Conference on Pattern Recognition*. Springer, pp. 72–81.

Bertiche, Hugo, Meysam Madadi, and Sergio Escalera (2019). "CLOTH3D: Clothed 3D Humans". In: *arXiv preprint arXiv:1912.02792*.

References II

Bogo, Federica et al. (2014). "FAUST: Dataset and evaluation for 3D mesh registration". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3794–3801.

Bronstein, Alexander M, Michael M Bronstein, and Ron Kimmel (2008). *Numerical geometry of non-rigid shapes*. Springer Science & Business Media.

Chen, Yang and Gérard Medioni (1992). "Object modelling by registration of multiple range images". In: *Image and vision computing* 10.3, pp. 145–155.

Chibane, Julian, Thiemo Alldieck, and Gerard Pons-Moll (2020). "Implicit functions in feature space for 3d shape reconstruction and completion". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6970–6981.

References III

Dai, Angela, Charles Ruizhongtai Qi, and Matthias Nießner (2017). "Shape completion using 3d-encoder-predictor cnns and shape synthesis". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5868–5877.

Davis, James et al. (2002). "Filling holes in complex surfaces using volumetric diffusion". In: *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*. IEEE, pp. 428–441.

De Aguiar, Edilson et al. (2008). *Performance capture from sparse multi-view video*. Vol. 27. 3. ACM.

Deng, Jia et al. (2009). "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Deprelle, Theo et al. (2019). "Learning elementary structures for 3d shape generation and matching". In: *arXiv preprint arXiv:1908.04725*.

References IV

Ganan, Stuart and D McClure (1985). "Bayesian image analysis: An application to single photon emission tomography". In: *Amer. Statist. Assoc*, pp. 12–18.

Gong, Ke et al. (2019). "Graphonomy: Universal human parsing via graph transfer learning". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7450–7459.

Groueix, Thibault et al. (2018). "3d-coded: 3d correspondences by deep deformation". In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 230–246.

Han, Xiaoguang et al. (2017). "High-resolution shape completion using deep neural networks for global structure and local geometry inference". In: *Proceedings of the IEEE international conference on computer vision*, pp. 85–93.

Hasler, Nils, Carsten Stoll, Bodo Rosenhahn, et al. (2009). "Estimating body shape of dressed humans". In: *Computers & Graphics* 33.3, pp. 211–216.

References V

Hasler, Nils, Carsten Stoll, Martin Sunkel, et al. (2009). “A statistical model of human pose and body shape”. In: *Computer Graphics Forum*. Vol. 28. 2. Wiley Online Library, pp. 337–346.

Hirshberg, David A et al. (2012). “Coregistration: Simultaneous alignment and modeling of articulated 3D shape”. In: *European conference on computer vision*. Springer, pp. 242–255.

Hu, Pengpeng et al. (2020). “Learning to Estimate the Body Shape Under Clothing from a Single 3D Scan”. In: *IEEE Transactions on Industrial Informatics*.

Jertec, Andrej et al. (2019). “On using PointNet Architecture for Human Body Segmentation”. In: *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, pp. 253–257.

References VI

Li, Chun-Liang et al. (2019). "LBS Autoencoder: Self-supervised Fitting of Articulated Meshes to Point Clouds". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11967–11976.

Litany, Or et al. (2018). "Deformable shape completion with graph convolutional autoencoders". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1886–1895.

Liu, Guilin et al. (2018). "Image Inpainting for Irregular Holes Using Partial Convolutions". In: *The European Conference on Computer Vision (ECCV)*.

Loper, Matthew et al. (2015). "SMPL: A skinned multi-person linear model". In: *ACM Transactions on Graphics (TOG)* 34.6, p. 248.

Ma, Qianli et al. (2019). "Dressing 3D Humans using a Conditional Mesh-VAE-GAN". In: *arXiv preprint arXiv:1907.13615*.

References VII

Marcard, Timo von et al. (2018). "Recovering accurate 3d human pose in the wild using imus and a moving camera". In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 601–617.

Marin, Riccardo et al. (July 2018). "FARM: Functional Automatic Registration Method for 3D Human Bodies". In: *arXiv e-prints*, arXiv:1807.10517, arXiv:1807.10517. arXiv: 1807.10517 [cs.CV].

Neophytou, Alexandros and Adrian Hilton (2014). "A layered model of human body and garment deformation". In: *2014 2nd International Conference on 3D Vision*. Vol. 1. IEEE, pp. 171–178.

Pishchulin, Leonid et al. (2017). "Building statistical shape spaces for 3d human modeling". In: *Pattern Recognition* 67, pp. 276–286.

Pons-Moll, Gerard et al. (2017). "ClothCap: Seamless 4D clothing capture and retargeting". In: *ACM Transactions on Graphics (TOG)* 36.4, p. 73.

References VIII

Robinette, Kathleen M, Hans Daanen, and Eric Paquet (1999). "The CAESAR project: a 3-D surface anthropometry survey". In: *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*. IEEE, pp. 380–386.

Saint, Alexandre, Eman Ahmed, et al. (2018). "3dbodytex: Textured 3d body dataset". In: *2018 International Conference on 3D Vision (3DV)*. IEEE, pp. 495–504.

Saint, Alexandre, Kseniya Cherenkova, et al. (2019). "BODYFITR: Robust automatic 3D human body fitting". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 484–488.

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, and Djamila Aouada (2020). "3DBooSTeR: 3D Body Shape and Texture Recovery". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

References IX

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, and Djamila Aouada (n.d.). "3DBodyTex: Building a textured dataset of 3D human scans with aligned body shapes". In: (). To be submitted.

Saint, Alexandre, Anis Kacem, Kseniya Cherenkova, Konstantinos Papadopoulos, et al. (2020). "SHARP 2020: The 1st Shape Recovery from Partial Textured 3D Scans Challenge Results". In: *Proceedings of the European Conference on Computer Vision (ECCV)*.

Saint, Alexandre, Abd El Rahman Shabayek, et al. (2017). "Towards automatic human body model fitting to a 3d scan". In: *Proceedings of 3DBODY. TECH 2017-8th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Montreal QC, Canada, 11-12 Oct. 2017*. Hometrica Consulting, pp. 274–280.

Szeliski, Richard and Stéphane Lavallée (1996). "Matching 3-D anatomical surfaces with non-rigid deformations using octree-splines". In: *International journal of computer vision* 18.2, pp. 171–186.

References X

Tiwari, Garvita et al. (2020). "Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing". In: *arXiv preprint arXiv:2007.11610*.

Vlasic, Daniel et al. (2008). "Articulated mesh animation from multi-view silhouettes". In: *ACM SIGGRAPH 2008 papers*, pp. 1–9.

Wu, Zhirong et al. (2015). "3d shapenets: A deep representation for volumetric shapes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920.

Wuhrer, Stefanie et al. (2014). "Estimation of human body shape and posture under clothing". In: *Computer Vision and Image Understanding* 127, pp. 31–42.

Xu, Zongyi, Qianni Zhang, and Shiyang Cheng (2017). "Multilevel active registration for kinect human body scans: from low quality to high quality". In: *Multimedia Systems*, pp. 1–14.

References XI

Yang, Jinlong et al. (2016). "Estimation of human body shape in motion with wide clothing". In: *European Conference on Computer Vision*. Springer, Cham, pp. 439–454.

Yang, Yipin et al. (2014). "Semantic parametric reshaping of human body models". In: *2014 2nd International Conference on 3D Vision*. Vol. 2. IEEE, pp. 41–48.

Yu, Tao et al. (2018). "Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7287–7296.

Zhang, Chao et al. (2017). "Detailed, accurate, human shape estimation from clothed 3D scan sequences". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4191–4200.

References XII

Zheng, Zerong et al. (2019). "Deephuman: 3d human reconstruction from a single image". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7739–7749.

Zuffi, Silvia and Michael J Black (2015). "The stitched puppet: A graphical model of 3d human shape and pose". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3537–3546.