BODYFITR: ROBUST AUTOMATIC 3D HUMAN BODY FITTING

Alexandre Saint ⋆   *Abd El Rahman Shabayek* ⋆   *Kseniya Cherenkova* ⋆†   *Gleb Gusev* †

Djamila Aouada ⋆   *Bjorn Ottersten* ⋆

⋆ SnT, University of Luxembourg   † Artec Europe SARL

ABSTRACT

This paper proposes BODYFITR, a fully automatic method to fit a human body model to static 3D scans with complex poses. Automatic and reliable 3D human body fitting is necessary for many applications related to healthcare, digital ergonomics, avatar creation and security, especially in industrial contexts for large-scale product design. Existing works either make prior assumptions on the pose, require manual annotation of the data or have difficulty handling complex poses. This work addresses these limitations by providing a novel automatic fitting pipeline with carefully integrated building blocks designed for a systematic and robust approach. It is validated on the 3DBodyTex dataset, with hundreds of high-quality 3D body scans, and shown to outperform prior works in static body pose and shape estimation, qualitatively and quantitatively. The method is also applied to the creation of realistic 3D avatars from the high-quality texture scans of 3DBodyTex, further demonstrating its capabilities.

Index Terms— 3DBodyTex, 3D body model fitting, automatic fitting, 3D body scanning, body landmark estimation

1. INTRODUCTION

Automatic fitting of human body models to 3D scans is an active research domain in computer vision and graphics due to its high industrial value for several applications including healthcare, digital ergonomic design, virtual changing rooms, security checkups, and different 3D entertainment applications [1, 2, 3]. Fitting the same model to different scans provides cross-parametrization and enables statistical shape analysis and skeleton/texture transfer. This includes dense point-to-point correspondence and applications such as symmetry detection, articulated subject matching, shape completion and pose reconstruction [4, 5, 6].

However, fitting a model to a 3D scan of a person in unconstrained scenarios is a complex task with considerable challenges in handling the body pose and shape variations, the variability in density and connectivity of 3D scan meshes and the accumulated noise from sensors and processing steps. For a reliable fitting, the model and scan have to be aligned.

Fig. 1: Sample fitting results of BODYFITR on challenging poses of 3DBodyTex [1] with scan (back), fitted model (front) and overlay (middle).

Previous works assume a standard body pose [7, 8] or the existence of guiding landmarks defined manually [9, 10]. The problem of automatically estimating consistent body landmarks has been tackled in the literature, but requires prior annotation of a dataset of 3D scans [11] or shows limited robustness to pose variations and arbitrary meshing [11, 12]. Recently, we have addressed these shortcomings in [13, 1] by proposing a fully automatic pipeline for body landmark estimation and 3D body model fitting (Fig 2a), which is more robust to pose variations and is independent of the meshing. It eliminates restrictive and tedious preprocessing routines by exploiting texture and widely-available annotations of body landmarks on 2D images. Despite the proposed improvements, some limitations remain, namely: 1) the estimation of 3D landmarks on the scan is done heuristically and may lead to unreliable predictions; 2) the landmark regressor on the body model is defined manually, introducing human bias; 3) the pose fitting considers all body parts together, hampering the convergence on complex poses; and 4) the shape fitting energy, based on a point-to-point metric, does not cope well with detailed features such as ears, nose, mouth and toes. These limitations affect the overall fitting quality.

This paper presents BODYFITR, a fitting method addressing the previously mentioned limitations and extending the method of [1], henceforth referred to as BODYFIT. BODYFITR makes the overall approach fully automatic without strong assumptions and provides a robust, systematic and accurate body model fitting to a scan in unconstrained sce-
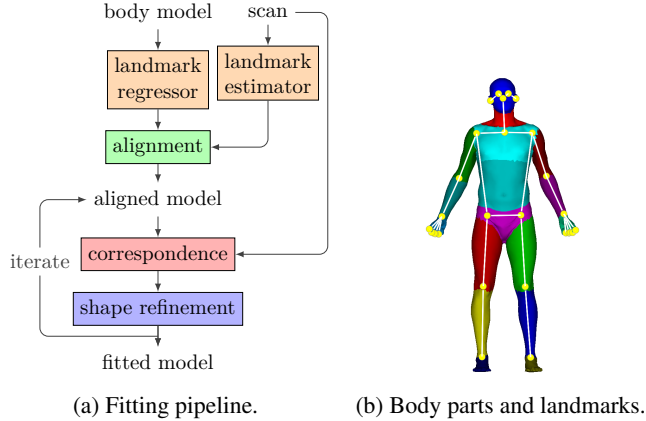
(a) Fitting pipeline. (b) Body parts and landmarks.

Fig. 2: Pipeline and body model.

narios. Our contributions to enhance the pipeline in Fig. 2a are summarised as follows: 1) a robust optimisation-based approach to automatically estimate 3D landmarks on a 3D scan; 2) an automatic and systematic approach to learn a landmark regressor on a 3D body model, where landmarks closely correspond to the landmarks detected on the 3D scan; 3) an incremental pose fitting strategy following the skeleton tree; 4) a robust energy function for shape fitting.

The rest of the paper is organised as follows: The background and motivation are given in Section 2. In Section 3, the proposed fitting method is presented. The experimental evaluation is given in Section 4, and the paper is concluded in Section 5.

2. BACKGROUND AND MOTIVATION

Fitting a human body model to a static 3D scan aims at estimating pose and shape parameters of the model such that the surfaces of the model and the scan match closely and are in correspondence. Our previous work, BODYFIT [1], follows the pipeline in Fig. 2a. The inputs are the body model and the target scan. The output is the fitted body model. There are three main stages: preprocessing, pose alignement, and shape refinement. During preprocessing, the respective landmarks of the body model and the scan are estimated. In the second stage, the landmarks serve as guides to align the pose of the body model with the scan. Finally, the shape of the model is refined iteratively to match the scan. Each iteration of the shape refinement is initialised by computing a dense vertex correspondence between the model and the scan. This correspondence is used in an iterative closest point (ICP) procedure [14] using the body model for regularisation. Both pose alignment and shape refinement steps are solved by minimising the following energy function,

$$E(y, r, s) = w_l E_l(y) + w_v E_v(y) + w_s E_s(s) + E_m(y, r, s),$$
(1)

with relative scalar weights w_*, and where y, r and s are the vertices, pose parameters, and shape parameters of the

body model, respectively. The energy terms in (1) ensure the consistency of the landmark positions (E_l), the close-fitting of the shape (E_v) and constrain the possible mesh deformations (E_m) and shape parameters (E_s), according to the body model. The shape fitting term,

$$E_v = \sum_{(y,z)\in C} \|y - z\|_2^2,$$
(2)

promotes the closeness of pairs of vertices, (y, z), belonging to the correspondence set between model and scan, C. The optimisation procedure and the formulation are further detailed in [1].

The body model used in BODYFIT to regularise the fitting is an adapted implementation of the Shape Completion and Animation of PEople (SCAPE) model [15]. Any other body model, such as [8, 16], could equivalently be used. The body model is defined on a template mesh with N vertices and a fixed connectivity pattern. For notation, the vertex positions are stacked as rows in the matrix $V \in \mathbb{R}^{N\times 3}$. A skeleton is defined on the template mesh, with a segmentation into body parts (Fig. 2b) linked by body joints. Moreover, $L = 26$ landmarks are defined over the body (Fig. 2b), divided into a subset of 13 body joints and 5 surface landmarks on the face and 4 surface landmarks on each hand. The face landmarks are the nose, eyes and ears. On the hands, the base knuckles are used. The position of landmark l, denoted $p_l \in \mathbb{R}^3$, is obtained by regressing the vertex positions of the body model with a vector $g_l \in \mathbb{R}^N$, such that,

$$p_l = V^T g_l.$$
(3)

In [1], the regressor, g_l, is defined manually.

Although BODYFIT is, to the best of our knowledge, the first fully automatic method to fit a body model to a 3D body scan, some components are limited in robustness or biased in their definition. Specifically, the landmarks are defined manually on the body model and estimated on the scan with a heuristic, while the fitting procedure may struggle with complex poses and the local shape features of realistic 3D body scan data. The next section details our proposed solutions to address these limitations.

3. PROPOSED ROBUST AUTOMATIC FITTING

This work brings notable improvements to the fitting pipeline in Fig. 2a by targeting each of its core modules. The following subsections detail our contributions to each of them.

3.1. Robust automatic 3D landmark estimation

To increase the robustness to outliers in a systematic way for the estimation of 3D landmark positions, we propose to replace the heuristic of BODYFIT with an optimisation-based formulation. As in [1], the 3D scan is projected onto 2D images from N_v viewpoints from which 2D landmark positions are estimated using OpenPose [17]. Because the 2D

projection parameters are known, the problem is solved in 3D space without reverting to backprojections on the generated images, as in [17]. To this end, in each view, v, 3D rays are reconstructed by joining the camera centre, $c \in \mathbb{R}^3$, to the detected landmark positions, $\{p_l \in \mathbb{R}^3\}_l$, using the known camera parameters. The rays are defined in parametric form, $r_l(\alpha) = c + \alpha_l d_l$, for some scalar parameters, $\alpha_l > 0$, and the direction, $d = p - c \in \mathbb{R}^3$, from camera to landmark. For each landmark, l, the goal is to find the best approximate intersection point of the rays, i.e. the point, $x_l^* \in \mathbb{R}^3$, minimising the distance to all rays. This is cast as the following l_1-regression problem,

$$x_l^* = \arg\min_{x,\alpha} \sum_{v \in \text{views}} \|x - (c^v + \alpha_l^v d_l^v)\|_1, \qquad (4)$$

with $\alpha \in \mathbb{R}^{N_v L}$, the vector of ray parameters, $\{\alpha_l^v\}_{v,l}$, one per view and per landmark. The l_1-norm makes the problem robust to outliers. Problem (4) is solved using iteratively reweighted least squares (IRLS) [18].

3.2. Automatic landmark regressor

To reduce bias and error in defining the landmark regressor on the body model, as with the manual process of [1], we propose to learn it from data, in a fully automatic and systematic way. The learning is based on reference landmark positions of the body model, in different configurations of pose and shape, generated synthetically using the automatic landmark estimation module of Section 3.1. The landmark positions are assumed independent of each other such that there is a separate regressor vector per landmark, $g_l \in \mathbb{R}^N$. As in Equation (3), the regression is performed separately in each spatial dimension, while the regressor vector is common to all.

The regressor is learned from K example data pairs, $\{(V^k, p_l^k)\}_k$, that represent the template mesh of the body model in different pose and shape configurations with the corresponding ground-truth landmark positions. Since the number of vertices is larger than the number of sample pairs, $N > K$, the system is underdetermined. For regularisation, sparsity is enforced on the regressor, with the added effect of encouraging it to rely on the most relevant vertices from the local neighbourhood. For landmark l, the problem is defined as a basis pursuit denoising (BPDN) [19] formulation,

$$\arg\min_{g_l} \frac{1}{2}\|b - Ag_l\|_2^2 + \lambda\|g_l\|_1, \qquad (5)$$

with, $A \in \mathbb{R}^{3K \times N}$, the stacked mesh vertices, $\{(V^k)^T\}_k$, and, $b \in \mathbb{R}^{3K}$, the stacked ground-truth landmark positions, $\{p^k\}_k$.

Following this procedure, the landmark regressor on the body model is learned in a systematic fashion and corresponds tightly to the landmarks estimated on a scan (Section 3.1), allowing a good initial alignment of the body model on the scan in complex shape and pose configurations.

3.3. Incremental pose fitting

When the pose of the scan is far from the initial pose of the body model, minimising (1) is an ill-posed problem. For regularisation, the problem is split into a series of simpler problems of gradually increasing complexity. In a first step, only the trunk of the body is aligned, using the corresponding landmarks. Then, the landmarks associated to the child body parts are added successively, following the tree structure in Fig. 2b, until all segments are finally considered together. This procedure has the advantage of starting each step in an initial position close to the optimum, reducing the search space for a more stable and faster convergence avoiding local minima.

3.4. Robust tight shape fitting

Our main objective is to tightly fit the shape in both smooth and detailed areas without any manual intervention. Instead of the point-to-point error metric (2), as used in [1], the fitting of corresponding model and scan vertices in the correspondence set, C, is performed with a point-to-plane error metric [20],

$$E_v = \sum_{(y,z) \in C} \|n_z^T(y - z)\|_2^2, \qquad (6)$$

with, $n_z \in \mathbb{R}^3$, the normal at point z on the target scan. As in Equation (2), this objective encourages pairs of corresponding vertices, (y, z), to be close. At the same time it gives more flexibility to the surface of the model to slide and adapt to the surface of the scan. This allows fitting curved and detailed areas robustly with less opportunity for stalling in local minima when the starting point is far from the optimum. Moreover, the point-to-plane metric makes the convergence faster, as shown in [14] for the related task of registering range images. Hence, our new formulation of the shape fitting improves on [1] on robustness and convergence.

4. EXPERIMENTS

The 3DBodyTex dataset [1] is used for validation of the proposed automatic fitting pipeline. 3DBodyTex contains 400 scans of 100 female and 100 male subjects. Each subject is dressed with light close-fitting clothing and performs two poses, a standard "U" or "A" pose and a further random pose. The SCAPE body model [15] is fitted to the 3DBodyTex data, using both BODYFIT [1] and the newly proposed BODY-FITR methods, as presented in Section 3. For a fair comparison, the evaluation setup is shared by both methods where possible. For the automatic landmark estimation on the scan (Section 3.1), the camera rotates around the centre of mass of the standing scan in $N_v = 18$ different views uniformly sampled in azimuth. To learn the landmark regressor of Section 3.2, a set of $K = 100$ pairs of example data is generated synthetically. The body model is put in varying realistic poses and shapes, and the corresponding ground-truth landmark positions are detected using the method in Section 3.1. During the pose and shape fitting, the weights of the energy functions

	BODYFIT	BODYFITR (proposed)
mean	3.10	2.40
median	1.54	1.33

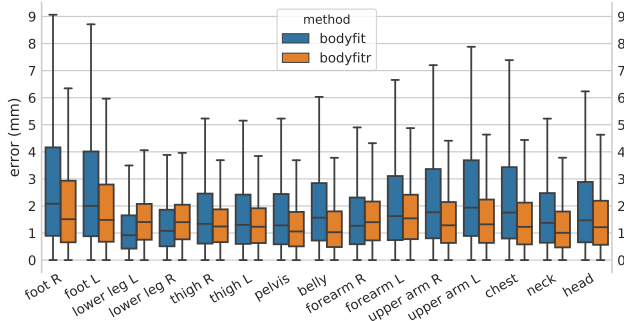**Table 1**: Overall fitting error (mm).

Fig. 3: Fitting error (mm) per body part on 3DBodyTex for BODYFIT and BODYFITR. Central bar: median. L: left, R: right.

in (1) are all initialised to 1. During shape fitting, the influence of the landmarks is gradually decreased by decaying w_l with a multiplicative factor of 0.5 at each iteration.

The fitting is evaluated quantitatively by estimating the distance between the fitted mesh and the target scan. The fitting error is defined for each vertex of the body model as the Euclidean distance to the corresponding closest vertex on the scan. Vertex errors are aggregated with a mean or median operator to obtain an overall measure across scans, for the whole mesh and for each body part separately.

On the full body, BODYFITR shows a lower overall fitting error, as reported in Table 1, indicating its better performance as compared to BODYFIT. It can be noted that for both approaches, the mean error is higher than the median. This suggests that most of the vertices are tightly fitted except for a smaller subset of outliers. This can be explained, for the most part, by limitations in the body model and realistic data. Indeed, the body model has clenched fists while most of the scans have open hands (Fig. 1); it smooths out sharper areas such as armpits, crotch, elbows and breast; and it discards non-skin regions including protruding hair and clothing.

The fitting error per body part is presented in Fig. 3. The error on the hands is not reported since the body model cannot fit open hands, as explained in the previous paragraph, making their analysis irrelevant. This is a limitation of the body model implementation and not of the proposed concept. It can be seen in Fig. 3 and Fig. 4 that most body parts have aggregated errors between 1 to 2 mm, for the proposed BODYFITR, and between 2 to 4 mm for BODYFIT. This is in accordance with the estimated error on the full body (Table 1). Moreover, the variability in the error is more controlled in BODYFITR, showing the stability and the robustness of the approach. Furthermore, the proposed method produces a lower fitting error on all body parts, except for the lower legs
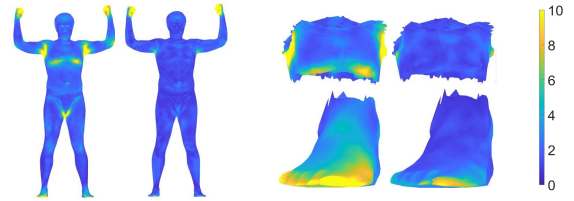

Fig. 4: Fitting error (mm) across all scans reported per vertex for BODYFIT (left) and BODYFITR (right) on the full body and with close-ups on the chest and right foot.

Fig. 5: Texture transfer from scan to fitted body model and examples of reposed model with texture (from left to right).

where the error is matched. This highlights the performance gain brought by BODYFITR. It can be noted that, although the right foot displays a higher error than all the other body parts with both methods (Fig. 4), the proposed approach performs better with a mean error of 2 mm, instead of 5 mm for BODYFIT. This comes from a specific challenging body pose in the data, where the subject, kneeling with the right leg, has a highly extended foot, see Fig. 1. This demonstrates the added robustness of BODYFITR in challenging poses.

Qualitatively, the results of BODYFITR can be observed in Fig. 1 for various challenging poses, with both male and female subjects. Moreover, as a concrete application, Fig. 5 shows how the fitting can be used to create a realistic avatar from the high-quality texture of 3DBodyTex. The texture of the scan is transferred to the fitted body model, which is animated in various new poses, keeping the fitted shape fixed.

5. CONCLUSION

We introduced BODYFITR, a novel robust approach to automatically fit a body model to a static 3D scan. The method builds on our previous work in [1] and improves the core fitting components. The proposed enhancements bring robustness to the pipeline and a systematic way of preparing it, removing the error and bias introduced by a human operator. An evaluation of the method conducted on the 3DBodyTex dataset [1] shows clear gains in performance with respect to [1], both quantitatively and qualitatively. The capability of the approach was also demonstrated for the creation of realistic 3D avatars. A limitation of the presented approach lies in the usage of a body model that cannot capture details and open hands. However, the approach is independent of the specific body model used. Hence, future works include augmenting the proposed fitting method with a body model incorporating finer body shape details. This will open the door to accurate and automatic estimation of the 3D body shape under a clothed static scan.

6. REFERENCES

[1] Alexandre Saint, Eman Ahmed, Abd El Rahman Shabayek, Kseniya Cherenkova, Gleb Gusev, Djamila Aouada, and Bjorn Ottersten, "3dbodytex: Textured 3d body dataset," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 495–504.

[2] T. Kwok, Y. Zhang, C. C. L. Wang, Y. Liu, and K. Tang, "Styling evolution for tight-fitting garments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 5, pp. 1580–1591, May 2016.

[3] Tan Xiaohui, Peng Xiaoyu, Liu Liwen, and Xia Qing, "Automatic human body feature extraction and personal size measurement," *ELSEVIER Journal of Visual Languages and Computing*, vol. 47, pp. 9–18, 2018.

[4] Oliver van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or, "A survey on shape correspondence," *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1681–1707, 2011.

[5] G. K. L. Tam, Z. Cheng, Y. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X. Sun, and P. L. Rosin, "Registration of 3d point clouds and meshes: A survey from rigid to nonrigid," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 7, pp. 1199–1217, July 2013.

[6] Liguo Jiang, Juntao Ye, Liming Sun, and Jituo Li, "Transferring and fitting fixed-sized garments onto bodies of various dimensions and postures," *ELSEVIER Computer-Aided Design*, vol. 106, pp. 30–42, 2019.

[7] Aggeliki Tsoli, Matthew Loper, and Michael J. Black, "Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses," *In Proceedings Winter Conference on Applications of Computer Vision (WACV)*, pp. 83–90, 2014.

[8] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black, "Smpl: A skinned multi-person linear model," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 248:1–248:16, Oct. 2015.

[9] Jinlong Yang, Jean-Sébastien Franco, Franck Hétroy-Wheeler, and Stefanie Wuhrer, "Estimation of Human Body Shape in Motion with Wide Clothing," in *ECCV 2016 - European Conference on Computer Vision 2016*, astian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, Eds., Amsterdam, Netherlands, Oct. 2016, vol. 9908 of *Lecture Notes in Computer Science*, pp. 439–454, Springer.

[10] Chao Zhang, Sergi Pujades, Michael J. Black, and Gerard Pons-Moll, "Detailed, accurate, human shape estimation from clothed 3d scan sequences," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[11] Christian Lovato, Umberto Castellani, Carlo Zancanaro, and Andrea Giachetti, "Automatic labelling of anatomical landmarks on 3d body scans," *Graphical Models*, vol. 76, no. 6, pp. 648–657, 2014.

[12] Stefanie Wuhrer, Chang Shu, and Pengcheng Xi, "Landmark-free posture invariant human shape correspondence," *The Visual Computer*, vol. 27, no. 9, pp. 843–852, 2011.

[13] Alexandre Saint, Abd El Rahman Shabayek, Djamila Aouada, Björn Ottersten, Kseniya Cherenkova, and Gleb Gusev, "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, 2017, pp. 274–280.

[14] Szymon Rusinkiewicz and Marc Levoy, "Efficient variants of the icp algorithm," in *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*. IEEE, 2001, pp. 145–152.

[15] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis, "Scape: shape completion and animation of people," in *ACM transactions on graphics (TOG)*. ACM, 2005, vol. 24, pp. 408–416.

[16] Hanbyul Joo, Tomas Simon, and Yaser Sheikh, "Total capture: A 3d deformation model for tracking faces, hands, and bodies," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[17] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *arxiv*, vol. abs/1812.08008, 2018.

[18] Paul W Holland and Roy E Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics-theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[19] Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[20] Yang Chen and Gérard Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.