Graphical Models 129 (2023) 101187

Contents lists available at ScienceDirect

Graphical Models

journal homepage: www.elsevier.com/locate/gmod

GIM3D plus: A labeled 3D dataset to design data-driven solutions for dressed humans

Pietro Musoni [a,*], Simone Melzi [b], Umberto Castellani [a]

[a] University of Verona, Strada le Grazie 15, 37134, Verona, Italy
[b] University of Milano-Bicocca, Viale Sarca 336, 20126, Milan, Italy

ARTICLE INFO

Keywords:
3D dataset
3D classification
3D segmentation
Clothed humans

ABSTRACT

Segmentation and classification of clothes in real 3D data are particularly challenging due to the extreme variation of their shapes, even among the same cloth category, induced by the underlying human subject. Several data-driven methods try to cope with this problem. Still, they must face the lack of available data to generalize to various real-world instances. For this reason, we present GIM3D plus (Garments In Motion 3D plus), a synthetic dataset of clothed 3D human characters in different poses. A physical simulation of clothes generates the over 5000 3D models in this dataset with different fabrics, sizes, and tightness, using animated human avatars representing different subjects in diverse poses. Our dataset comprises single meshes created to simulate 3D scans, with labels for the separate clothes and the visible body parts. We also provide an evaluation of the use of GIM3D plus as a training set on garment segmentation and classification tasks using state-of-the-art data-driven methods for both meshes and point clouds.

1. Introduction

In the last decades, the need for synthetic world representation has grown exponentially, especially for virtual and augmented reality environments. There are always more application sectors that require effective, realistic simulation and animation of digital objects and scenes for the purpose of entertainment, training, education, and so on. Human bodies are largely analyzed shapes as the natural target of interactions of the user. The modeling of human shapes is a complex task due to the strong variation of body characteristics (e.g., male or female, fat or thin, small or tall) and the presence of non-rigid deformations in changing the posture or facial expression. This modeling task becomes more challenging when the human is wearing clothes, especially for dynamic scenarios. In particular, it is important to capture the garment deformations and wrinkles that depend on the cloth materials (e.g., elastic properties) and the interaction between the garment and the human body (e.g., human shape and pose). The standard approach is based on an approximation of physically-based simulations [1–3] adopting a fully synthetic paradigm. To this aim, several software and tools are available that require the work of expert artists (e.g., Marvelous Design[1]). Another promising approach consists of modeling the dressed human from the observation of real samples using body scanner technologies. This paradigm brings several advantages, such as the improvement of the realism of the obtained digital representations, the automation of the process, and the accuracy in capturing local geometric details. On the other hand, the acquired shapes are characterized by noise, holes, and irregular tessellation. Moreover, this representation lacks knowledge of the semantic components among the garments and the human and between the garment parts. Therefore, a labor-intensive activity is still demanded to make the acquired shape ready for modeling and animation. In order to improve the automation of the modeling-from-real-scans approach, a crucial step consists of exploiting a 3D shape segmentation strategy to detect the garment types and separate them from the underlying human body. These tasks are particularly challenging for dressed humans due to the large variability of shapes (global) and the arising of random wrinkles (local) on the garments. A recent trend consists of adopting learning-based technologies that exploit data-driven paradigms to cope with these challenging tasks. In particular, new neural network architectures have been proposed for feature extraction on 3D shapes represented by unstructured 3D point clouds [4,5] or triangular meshes [6]. The availability of an appropriate dataset of labeled samples is a crucial need in order to learn how to separate the human body from the worn garments (segmentation) and how to categorize each typology of cloth (classification) within a unique mixed data as a 3D scan, to extend this learning methodology to dressed humans.

* Corresponding author.
E-mail address: pietro.musoni@univr.it (P. Musoni).

[1] https://www.marvelousdesigner.com/

https://doi.org/10.1016/j.gmod.2023.101187
Received 24 February 2023; Received in revised form 2 July 2023; Accepted 6 July 2023
Available online 4 August 2023

Fig. 1. Some examples of the clothes and the subjects composing the dataset. The 3D meshes of GIM3D plus can have an upper cloth (in blue) and a bottom cloth (in yellow) with different styles (t-shirts, singlets, long-sleeved shirts) and fabrics or a unique garment: the jumpsuit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this paper, we proposed GIM3D plus, a novel dataset of labeled 3D shapes for dressed human segmentation and garment classification. Our GIM3D plus is composed of several human subjects with specific geometric characteristics. Each subject is wearing different garment types, i.e., t-shirts, long shirts, tops, shorts, trousers, skirts, and jumpsuits. All these clothes appeared in different sizes. Moreover, the human subjects are in different poses allowing the observation of a large variety of cloth wrinkles caused by human motion. The GIM3D plus dataset has been created from a selection of digital garments available on the CLOTH3D [7]. A well-designed pipeline has been implemented to merge the different layers composed of the human body and related clothes in a unified mesh with the aim of emulating some characteristics of the output of a body scanner. In this fashion, GIM3D plus provides a single watertight mesh for each sample with a label for each vertex to identify the human body or the garment types. Other datasets are available for dressed-human [7–10], but they have not been specifically designed for human-cloth segmentation and garment classification. For instance, in [9], the BUFF dataset has been proposed to estimate the human shape under clothes; therefore, cloth labeling is unavailable. In [7,10], the CLOTH3D and DeepFashion datasets provide the garments in separated meshes for the estimation of 3D cloth from a single image. In [8], the SIZER dataset has been introduced to learn the generation of synthetic garments of different sizes (some examples of the variability of the dataset can be observed in Fig. 1). Our main contribution is threefold:

- we provide the first dataset of labeled dressed humans that has been specifically designed for dressed human segmentation and garment classification. The dataset is publicly available in.[2]
- we show that our dataset is largely expressive to enable a neural network to learn the segmentation and classification tasks reliably. Three state-of-the-art neural network architectures for 3D segmentation have been evaluated, namely PointNet [4], PointNet++ [5], and DiffusionNet [6].
- we show that the shapes of our dataset correctly emulate the most peculiar characteristics of the output of a body scanner as reported by our experiments on a separate test-set of real dressed body scans.

The rest of the paper is organized as follows. In Section 2, we describe state of the art on publicly available datasets of clothed 3D. In Section 3, we introduce our GIM3D plus dataset and the working pipeline to generate it from other sources. In Section 4, we present exhaustive experiments on the evaluation of our dataset in training different neural networks for garment segmentation tasks. Finally, in Section 5, conclusions are drawn, and future works are envisaged.

This article is an extended version of "GIM3D: A 3D dataset for garment segmentation" [11]. In this work, we expand the applications we can address with GIM3D by adding to the dataset proposed in [11] all the necessary information to target the classification task. For each cloth in GIM3D plus, we provide the pointwise labels dividing the outfits of each human character into new, specific garment classes (see Fig. 4). In addition, we expand the dataset with over 400 more shapes with skirts as a lower garment. Furthermore, we perform new experiments on garments classification (Section 4.2), providing the numerical evaluation of the classification into these new classes of clothes (Table 3). Finally, we analyze the issues in the classification of some challenging instances in our dataset by comparing recent data-driven methods (see Section 5 and Fig. 8).

2. Related works

The main motivation for our work is the lack of available data for 3D garment segmentation and classification, which are real challenges for the computer graphics community. Some really interesting datasets are shared by the community built for different tasks, such as human

[2] https://github.com/PietroMsn/GIM3D

body registration, 3D reconstruction of clothes from 2D images, and 3D garment generation. In the following paragraphs, we present the main works for this typology of data, the datasets exposed here are available, under request, on the websites of the projects.

Sizer [8] is composed by 3D scans of 100 subjects with 10 different garments classes. The authors provide also the segmentation of the garments and the registrations of the scans with *SMPL+G* [12], a parametric 3D model of clothed characters. The aim of the work is to predict the clothing over an avatar in the function of the size of the cloth. They develop two separate network architectures. The ParserNet separates a single registered mesh into a multi-layered representation of the body and the clothes, the SizerNet predicts the garment in the function of the cloth. This dataset has a large variety of subjects and clothes, they provide also the labeling of the different garments which is a rare feature in this type of dataset, but all the subjects are acquired in the rest pose and the lack of variability of poses can be a limitation in the generalization of the data for certain tasks.

Bodies Under Flowing Fashion (BUFF) [9] is a rare example of 4D captured scans with clothes, the majority of data available acquired from the real world are in a single pose. Here the authors provide sequences of 6 different subjects in two different outfit styles (t-shirt with trousers and long-sleeved shirt with shorts). For each subject and each clothing category, the dataset contains 3 different motions with a length of between 4 and 9 s (200–500 frames) with a total of 13,632 scans. The aim of the work is the estimation of the human body under the clothes using the SMPL [13] parametric model. The scans contained in the dataset are in different poses, which is rare for real scans datasets, especially for clothed models. The total amount of 3D models is very large but the variability inside the data is limited, due to the difficulties of the scan acquisitions, so for each sequence, hundreds of 3D models are given, but the difference between the poses is very low. For deep learning purposes this can be a limit since the scans are very similar in the sequences of the same motion of the same subject. The clothing labels are not provided so the dataset is not suitable in the training process for segmentation and classification tasks.

DeepFashion 3D Deep Fashion3D [10] contains over 2000 models acquired from real clothes in different clothes and covers 10 different garment categories. The work is made to provide a large amount of data for 3D model reconstruction from single-view 2D images. The data include for each 3D model a point cloud of the reconstructed cloth, the multi-view images used for the reconstruction, the pose of the estimated 3D human model under the garment and the feature lines, annotations over the point cloud which highlight the junctures over the cloth surface. This dataset provides a large amount of 3D models of garments from real deformed objects and this typology of data is very scarce compared to undressed human models but they do not provide complete outfits, each garment has its own 3D human character.

CLOTH3D is a synthetic dataset composed of physically simulated clothed models. The dataset [7] provides a huge amount of 3D clothes in motion, they use physical simulation to deform the surface of the garments in a realistic way over the motion of large variability of 3D avatars. The characters of this dataset are created through the use of SMPL model. With this morphable model is possible to create a large variety of human models by varying the 10 shape parameters. With the CMU Mocap dataset, they were able to obtain a large number of animations (each sequence provided has 300 frames), and for each subject and each motion the authors provide the parameters of the human model, the template of the single garment meshes, the point cloud of the clothes for each frame, the category of the outfit and the fabric (the main parameter used for the physical simulation). The sequences of this dataset are over 7000, and for each one 300 frames are provided so it is very suitable for tasks that require a large amount of data, such as training a network. The data, though, are synthetic and the very regular tessellation (required for a good physical simulation) is not suitable for tasks involving noisy data with very irregular tessellation, such as 3D scans.

All these works have peculiarities very useful for the different tasks in which they are involved. In addition, in recent years, a large variety of new works, such as [14,15], provide methods for the acquisition of 3D human models with common devices but similar to other methods requires additional human interaction for annotation for segmentation and classification tasks, which becomes very tedious for building large datasets. For these reasons we decided to build our own dataset for our target, we use the huge variety of clothes and the realistic dynamics of the garments in CLOTH3D to build our dataset GIM3D plus. In the following section, we describe the composition of our dataset, the steps to obtain the data that we wanted to manage, and the choices in building it. A first version of the GIM3D dataset was presented in [11]. In the previous version of GIM3D the 3D shapes were divided into three segments: body, upper and lower clothes. In GIM3D plus we extend the information on each shape by dividing each garment in classes according to the different typologies of clothes (more details in Section 3). The upgraded version of the dataset include also a new type of outfit, comprehending the skirts, and increasing the number of unique shapes in the dataset. In Section 4 we extend the experiments with respect to the previous version with an analysis on garment classification performed by data-driven methods trained on our dataset, by using the new information included in GIM3D plus.

3. 3D Garments In Motion (GIM3D plus) dataset

In the following paragraphs, we present all the details of the composition of our dataset and the passages to obtain the data.

Data source. To build our dataset, we use the physical simulations contained in the CLOTH3D dataset. In this work, the authors created several sequences of animation using the parametric human model SMPL and the CMU Mocap dataset[3] of animations. With the use of these 3D avatars they were able to create several clothes simulations, each subject has a sequence of 300 frames. For the physical simulation a large variety of fabrics are represented, by using different set of parameters of the mass–spring model used for the simulation. The simulations are computed by using Blender[4] that gives several fabrics presets (*cotton*, *leather*, *silk* and *denim*). The data provided by CLOTH3D are the separate garments resulting from the physical simulation and the parameters of the related 3D character.

Data processing. For our dataset, we merged in unique meshes the avatar and the separate garments in order to create a synthetic scan. We create the final 3D models of our dataset in three separate steps: we close the holes in each separate cloth mesh, then we obtain a watertight manifold surface and finally, we simplify the mesh to obtain a more suitable resolution of vertices for our task. These preprocessing steps are implemented as follows:

- We take the separate mesh of the garments of CLOTH3D and we close the holes of the model using *meshfix* [16]. This step helps the following step to create a watertight closed mesh.
- The method presented in [17] is a robust software to produce watertight manifolds from triangle soups This step helps us to empty the vertices inside the surface of our 3D models. The output mesh is very dense with hundreds of thousands of vertices (as we can see in Fig. 2) (b).
- The final step is mesh decimation, we use the *Quadratic decimation* implemented in *pyvista* library and we obtain meshes with around 20k vertices. The decimation process does not remove the details of the model. We can evaluate it by comparing (a) and (c) of Fig. 2

[3] www.mocap.cs.cmu.edu/

[4] https://www.blender.org/

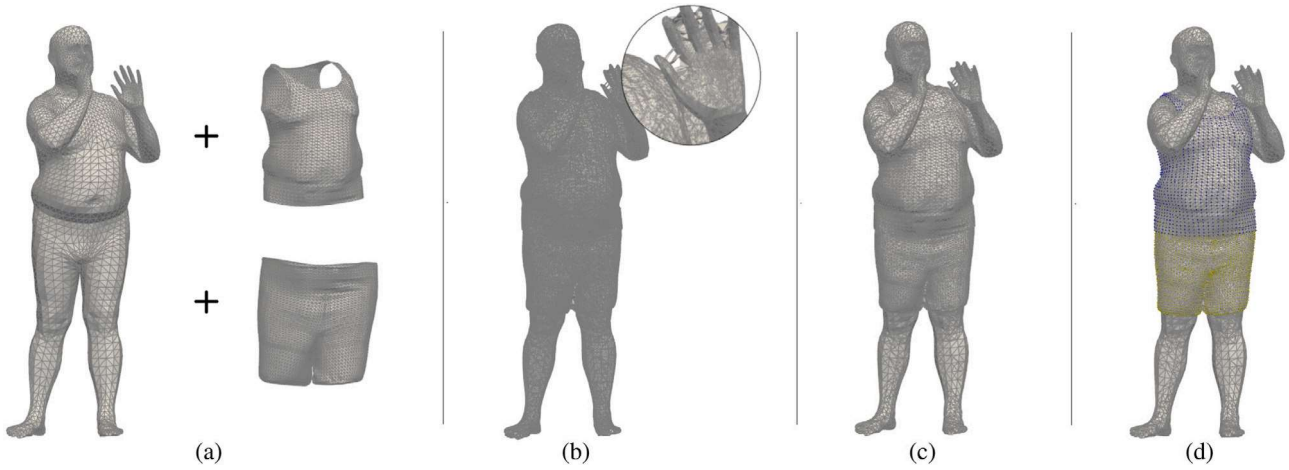

Fig. 2. A visualization of the proposed pipeline to produce our dataset. In (a) the 3 separate 3D models coming from the CLOTH3D dataset, then (b) we merge the body with the clothes and we obtain a very dense watertight mesh. In (c) lower the number of vertices maintaining the details. In the last step (d) we use the separate meshes of the clothes to apply the labels on the mesh using the nearest neighbor algorithm.

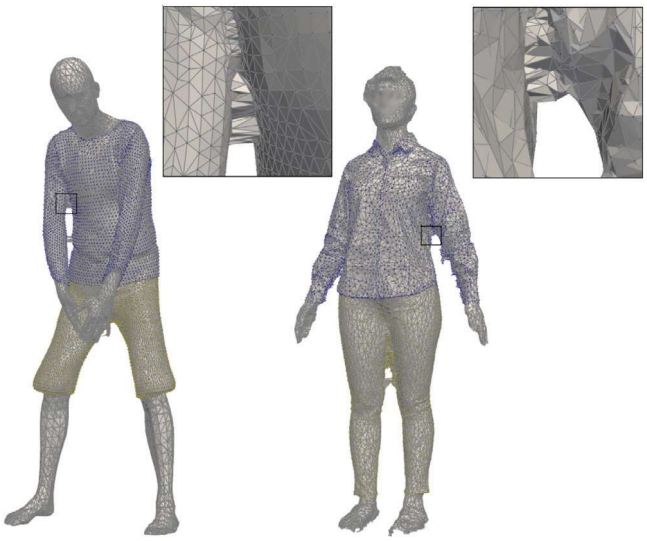

Fig. 3. An example of artifacts on the dataset shapes. On the left, we have a synthetic model from our dataset, and on the right, a real 3D scan from the SIZER dataset. We highlight how the process of obtaining a single manifold mesh produces artifacts similar to the output meshes of 3D scanners.

The second step of this process creates some artifacts, especially between the fingers and along the arms (in the poses near the body) and the same artifacts can be found in 3D scans (Fig. 3). At this point, we have a unique mesh composed of the subject and the outfit and we want to store the information of the label of each point of the 3D model.

Data labeling The main goal of this work is the creation of a new dataset for the supervised garment segmentation task. To this aim, the most critical aspect, together with the variability of the involved shapes, is the labeling. In order to transfer the semantic information that is available from the source data to the output mesh, we use the separate garment meshes provided by CLOTH3D. All the preprocessing steps described above do not change the position in the 3D space of the single meshes and add only a few vertices that do not deform the starting surface. Given that, we first create a denser version of the starting regular mesh of each cloth, and then through the nearest neighbor algorithm we label every single point of the unique mesh as upper or lower cloth. The remaining points are marked as body. Some points, though, do not overlap precisely with the starting meshes since they are created in the last step of the merging process. For this reason, we add a final adjustment step by using a voting method to decide whether a vertex is correctly labeled as body or not. If the majority of the nearest points, within a fixed radius (0.001 m), are labeled differently, we change the label of the examined point. We found out that this method works very well since the wrongly labeled points are very sparse on the clothes surfaces.

In GIM3D plus, together with the labels for the body, the upper and the lower clothes, we provide the labels of the different typologies of garments. We divided by hand all the models in our dataset in the following clothes classes: *t-shirts*, *shirts*, *shorts*, *trousers*, *tops*. An example for each class can be seen in Fig. 4. The combinations of these typologies of garments compose six different outfits: *t-shirt* + *shorts*, *t-shirt* + *trousers*, *shirt* + *shorts*, *shirt* + *trousers*, *top* + *shorts*, *top* + *trousers*. In Section 4, we provide the results of the garments classification task in these different clothes. In addition, we include a sixth class for the experiments: *the skirts*. The outfit including this type of garment are *t-shirt* + *skirt*, *shirt* + *skirt* and *top* + *skirt*.

Data Description. The dataset is composed of a total of 4623 meshes, divided into 1851 with two separate outfits (shirts and pants) and 2772 with a single garment (jumpsuit). We decided to use three typologies of outfits: (i) shirts and pants, (ii) top and pants, and (iii) jumpsuit. In Fig. 5 we can see the three categories of the dataset. Despite the name of these categories, the shapes of the different clothes composing every outfit differ in many ways: (i) the length of the sleeves (in the t-shirts category both short and long sleeves are included and the same for the trousers), (ii) the tightness (different garments give very different fit, also depending on the subject involved), and (iii) the fabric (the same garment may produce different wrinkles during the physical simulation, it depends by the fabric parameters). Some examples are shown in Fig. 5. All the meshes in the dataset have around 20k vertices. The SMPL model used for the avatars has a unique template of 6890 vertices, but each garment has a different resolution and the visible body parts of the underlying body vary for each garment.

For our GIM3D plus dataset, we took a single random pose available from the motion sequences. We will publicly share the code that implements our procedure to create the output meshes from the source data. This gives, in principle, the possibility to enlarge the number of data using all the 300 frames of the motion sequences. Of course, the difference of the pose between a frame and the successor is limited but even taking 3 frames from each sequence gives the opportunity to triple the number of shapes in the dataset. In Section 4 we test each subset composing our dataset. For the classification experiments, we provide an extra set of data with skirts as the lower part of the outfit and the shirts as the upper part. The shapes with skirts are 410 and bring the total number of clothed models to 5033.

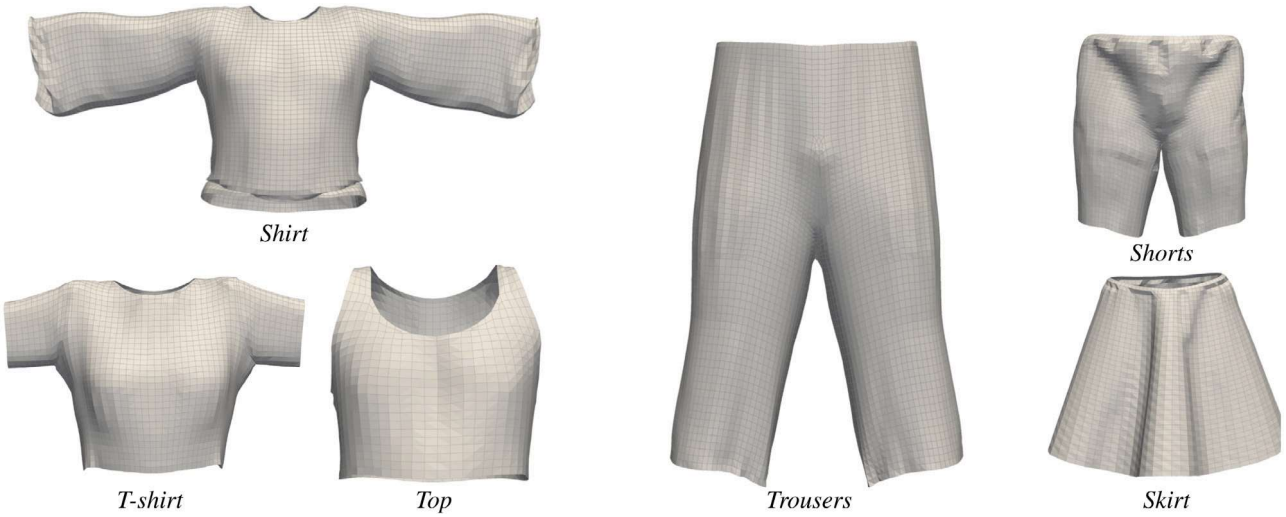

Fig. 4. The six different classes of clothes are t-shirt, shirt, trousers, shorts, top and skirt. Within the same class the proportions differ and the variability of each garment is very large, in the labeling we made specific choices. In particular, we classify as t-shirts if the sleeves are above the elbow and in the same way we classify as shorts all the pants above the knees.

Table 1
IoU(%) results on clothes and body segmentation with PointNet, PointNet++ and DiffusionNet architectures. In the first column, we put the training set used for each experiment, and in the second column the sets of data on which we test the methods. For the training set, we use *upper (shirt) + lower* which is composed of sleeved shirts, *upper (top) + lower* has only tanktops and tops without shoulder straps for the upper clothes. The *full dataset* is a merge of the first two datasets and the *binary labels dataset* is composed of 3D characters wearing jumpsuits, so the segmentation, in this case, involves only two categories: cloth and body. For the first set of experiments, we use a portion of our dataset. In the following two, we use a subsection of BUFF and SIZER datasets.

Training set	Test set	PointNet	PointNet++	DiffusionNet
upper (shirt) + lower	upper (shirt) + lower	82.61	85.41	**93.33**
upper (top) + lower	upper (top) + lower	**86.12**	**89.46**	92.24
full dataset	full dataset	82.85	87.80	89.27
binary labels dataset	binary labels dataset	90.02	94.02	90.84
upper (shirt) + lower	SIZER real scans	**70.59**	71.86	**74.31**
upper (top) + lower	SIZER real scans	65.37	65.00	69.59
full dataset	SIZER real scans	69.87	**72.61**	72.07
upper (shirt) + lower	BUFF real scans	72.68	74.91	**76.68**
upper (top) + lower	BUFF real scans	71.05	73.80	76.25
full dataset	BUFF real scans	**75.10**	**75.41**	74.54

Table 2
IoU results on clothes and body segmentation with PointNet, PointNet++ and DiffusionNet architectures. In this experiment, we consider two settings: we test the networks trained with upper (top) + lower on shapes with upper (shirt) + lower and vice versa.

Training set	PointNet	PointNet++	DiffusionNet
upper (shirt) + lower	85.40	**88.21**	87.17
upper (top) + lower	73.93	**75.69**	75.01

4. Experiments

We tested our GIM3D plus dataset on state-of-the-art methods for dressed human segmentation and garment classification. We evaluated the results using both the entire dataset and its different subsets. For surface segmentation and garments classification tasks we use the three following deep learning architectures: (i) *PointNet* [4], (ii) *PointNet++* [5] and (iii) *DiffusionNet* [6]. *DiffusionNet* operates on polygonal mesh, the two other architectures take as input only Point Clouds. For all the three networks we use the basic implementation provided by the authors and adapted it for our dataset.

4.1. Dressed human segmentation

As shown in Table 1 we followed four different training scenarios for the dressed human segmentation, we train the three different networks using:

- the *upper (shirt) + lower dataset*, composed of 815 meshes in shirt and pants outfit,
- the *upper (top) + lower dataset*, which has 1036 subjects with singlets or tops with pants,
- the *full dataset*, that is the fusion of the two previous datasets (with 1851 3D shapes), and
- the *jumpsuit dataset*, which is composed of only one garment for the whole body and can be seen as the fusion of pants and shirt. We use this data for testing the garment-body binary segmentation and it is referred to as *binary labels dataset*.

We test the three methods with a different set of data. In the first part of segmentation experiments (the first four rows of Table 1), we use as test set the same outfit categories of the training set and the models are taken from our dataset. In the second and third sets of segmentation experiments we test the three methods on shapes from

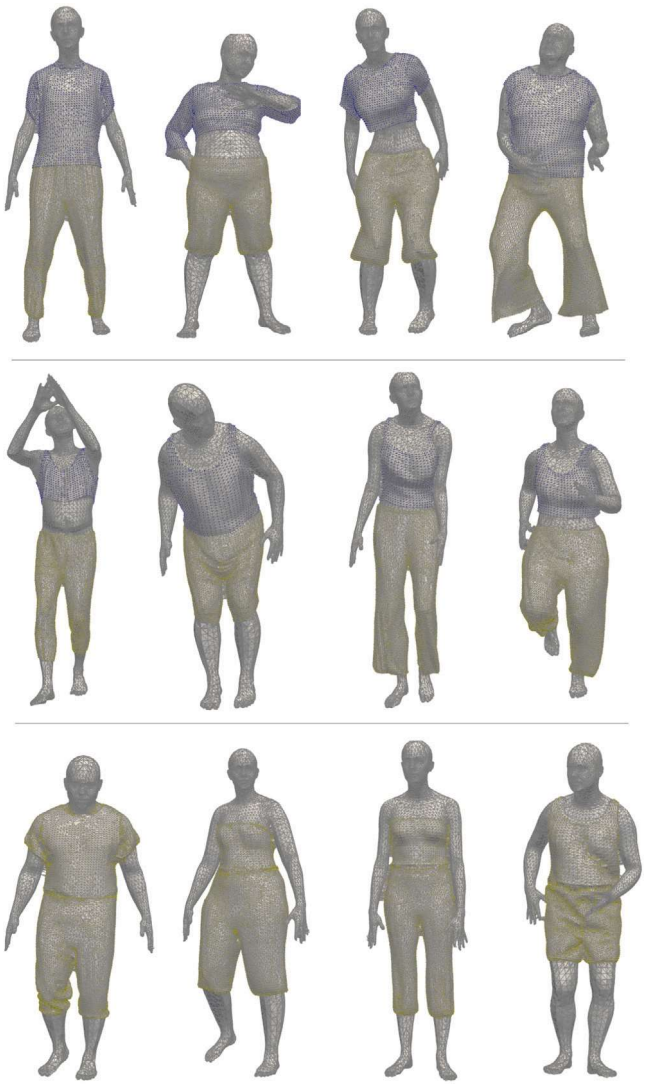

Fig. 5. Some examples of the three subsets of our dataset. At the top some samples of the upper (shirt) + lower dataset, in the middle the upper (top) + lower dataset, and at the bottom, the jumpsuits. Here we can see that even in the same category the variability of the shapes is considerable. Under the shirt label we can see several lengths of sleeves, in the top category in some cases there are shoulder straps in some cases none and the lengths of the pants of each category vary from over the knees to the ankles.

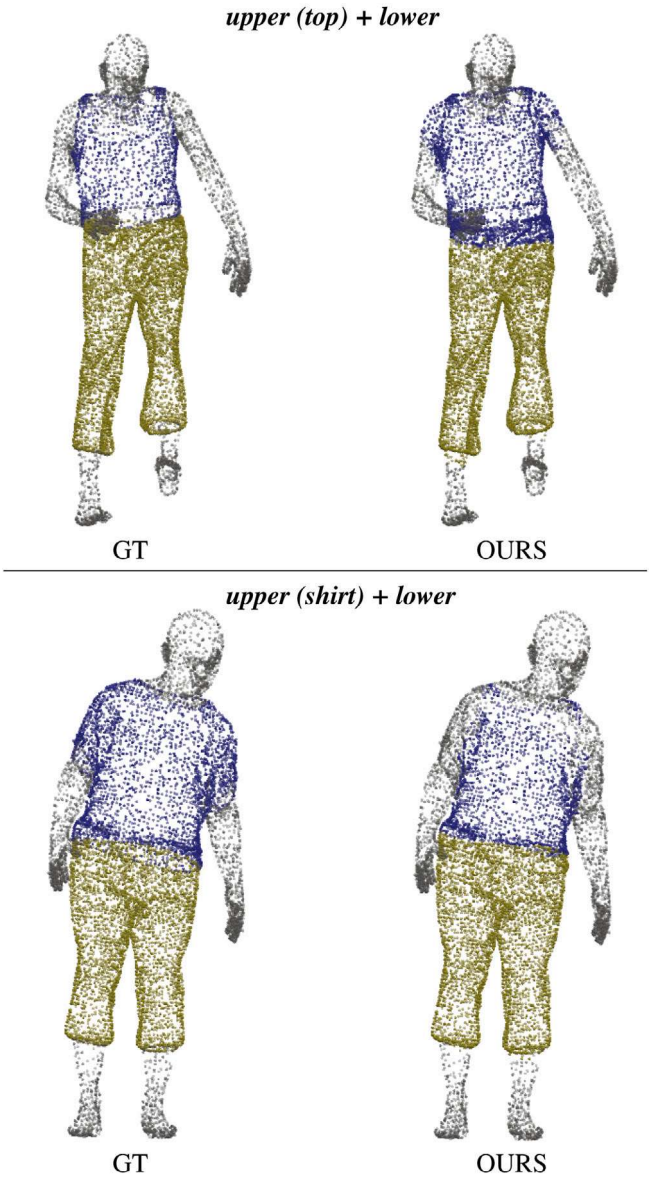

Fig. 6. A qualitative example of segmentation from Table 2. We report the ground truth on the left and, on the right, our result. In the first row, we have an example of segmentation where the training set is the *upper (shirt)* + *lower* collection and the test set is the *upper (top)* + *lower* one, in the second row we switch the training and the test sets. These examples are both results of *PointNet++* method, which provides the best performance.

SIZER and BUFF datasets. In Table 2, we make a cross experiment between the two different outfits, so we test the *upper (top)* + *lower* data on the architectures trained with *upper (shirt)* + *lower* dataset and vice versa. We can see that the methods perform better in the first case so we can observe that the networks can generalize better when trained on shirts and they can cope with the missing parts (i.e. the sleeves) of the tops in the test set. Some qualitative results of this experiment can be seen in Fig. 6. The evaluation metric for the segmentation task is the IoU (Intersection over Union) on points for each shape, as described in the original paper of *PointNet*. Then we average the results over all the shapes. We use 80/20 train-test split for all the experiments (e.g. for the full dataset the training and the test sets are composed of 1480 and 371 shapes respectively). In Fig. 5 some examples of the three different portions of our dataset. As can be seen, the two first have three different labels: the body, the upper cloth, and the lower cloth. The *jumpsuit dataset* has only two labels, one for the cloth and one for the body parts. We tested the methods mentioned above also on a small set of real 3D scans, taken from SIZER and BUFF datasets. As mentioned in Section 3, the SIZER dataset has a large number of subjects, but all of them are in A-pose. Meanwhile, the BUFF dataset has 3D scans in different poses but a limited number of subjects. We selected 15 shapes from each of these datasets and we tested the three networks trained on our dataset. The outfits of these shapes vary in the length of pants and shirts. The labels for this small test set have been manually defined.

In Fig. 7 some qualitative results from these experiments and in Table 1, the quantitative results of the networks on these two test sets is shown. The *full dataset* has inferior results in relation to the subsets, which are composed of and tested individually. Merging the two categories introduces a huge variability in the possible shapes of the garments. We can see that, for the 3D scans datasets, the best training set for the segmentation task is the *upper (shirt)* + *lower*, since the 30 shapes are in huge majority composed of sleeved shirts and just a couple of subjects wear the top category garments. The tests with SIZER dataset perform a bit lower than BUFF, the meshes in SIZER contain

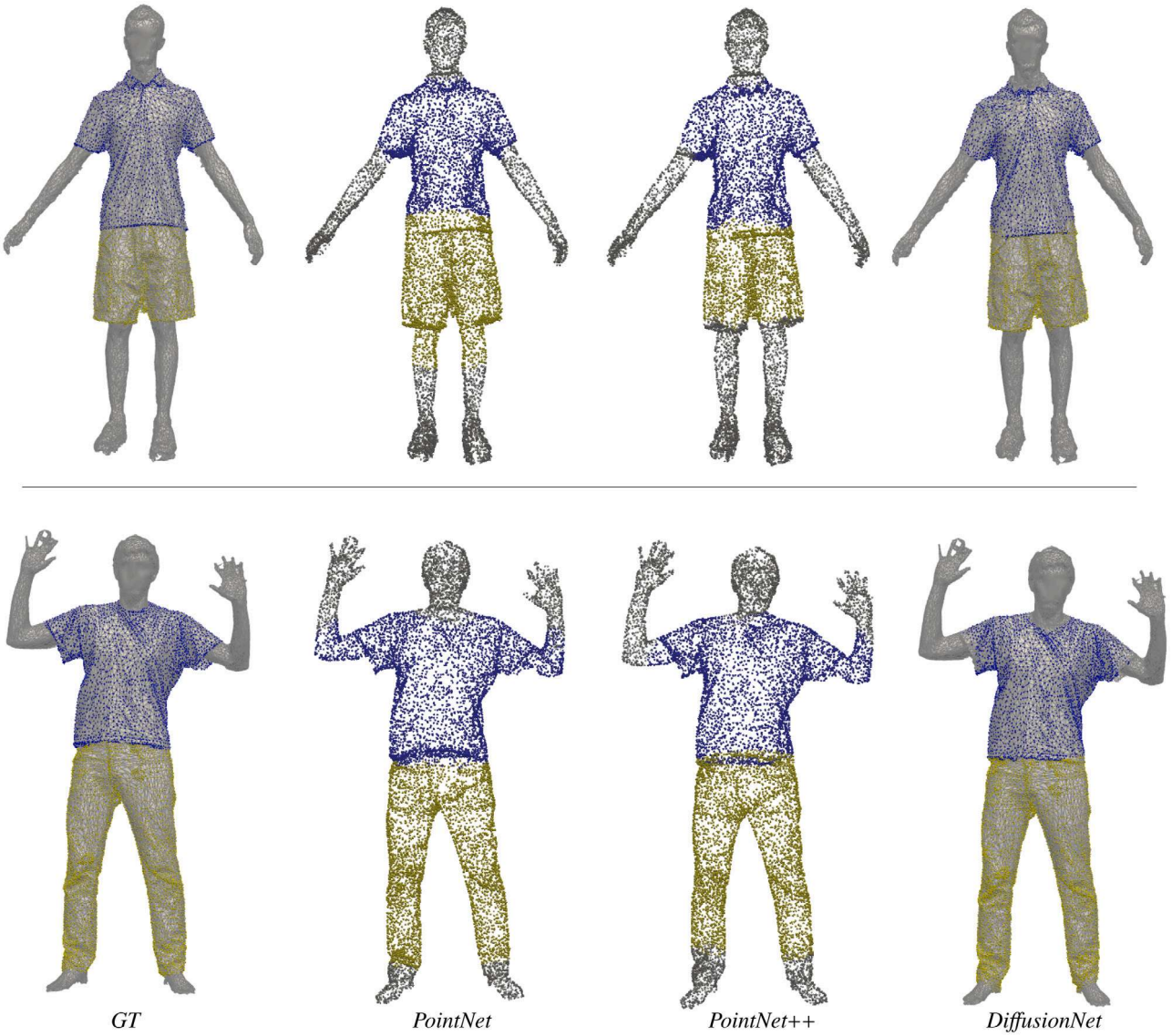

Fig. 7. Two examples of real scans from BUFF and SIZER datasets labeled using the three networks we consider in our experiments and exploiting our dataset as the training set. In the first row, a scan from SIZER (every scan in SIZER is in A-pose). In the second row, a scan from BUFF in an arbitrary pose. The first mesh on the left is the Ground Truth from our dataset, then in order the results using, *PointNet*, *PointNet++* and *DiffusionNet*. The two first methods operate on Point clouds, *DiffusionNet*, meanwhile, works with meshes.

more noise in comparison with the ones in BUFF. Another characteristic of SIZER is that the subjects wear shoes, all the other 3D models and, above all, the training set meshes have naked feet. Nevertheless, the points on the feet are generally labeled correctly as body. In general, we can see that *PoinNet++* is the method that performs better on point clouds (as we can see in the two central models of Fig. 7, especially on the sleeves and on the legs) and *DiffusionNet* is the method that performs better in overall. Our dataset is composed of manifold meshes and *DiffusionNet* performs better with meshes (despite can also take pointcloud as input). *PoinNet++* also uses the information of the points normals, concerning the *PoinNet* architecture, and in the case of meshes they can easily be estimated. For these reasons, these methods perform better than the basic *PoinNet* method. As exposed in Table 1.

4.2. Garment classification

The second set of experiments performed on GIM3D plus is related to the garment classification task. As seen before in the segmentation tests, the possible variations on the typology of garments for both the upper and the lower part of the outfit are very wide. Some examples are visible in Fig. 5, as can be seen, the *upper (shirt) + lower* subset contains outfits with both sleeves and pants with very different lengths. We can easily understand that a dataset like GIM3D plus should help in data-driven methods for the classification of garments. For this reason, we provide some experimental results by using the same Network architectures described in the previous section on the classification of clothes of dressed humans. The 5 different classes of garments we use for the classification are the same exposed in Section 3, which create the 6 possible outfits inside GIM3D plus. In Table 3, we can see the overall accuracy of the classification of the different clothes for all the test sets that we use for the segmentation experiment. As in the previous experiments, we can see that *DiffusionNet* performs better in most of the cases. In addition to the five classes of garments, we decide to perform an ulterior experiment on another challenging class: the skirt. The amount of 3D models with a skirt in the outfit used for this experiment is 410, and the classes for the upper clothes are the same that we previously exposed. We can see that in this set of tests, the accuracy is lower than the previous ones, as we can expect, but due to the very different conformation of this new class we consider the

Table 3
Classification results on the three test sets (one for each row). The table is divided into two parts with respect to the adopted training set. In the first three rows, we report the overall accuracy results with the full dataset without skirts. In the last three rows, we expose the same experiments performed, selecting as training data the full GIM3D plus dataset plus more than 400 shapes with a skirt in the outfit. As for the segmentation task, we compare three different feature extractors (one for each column) and we select the test set as a portion of GIM3D plus, 15 shapes from SIZER and 15 from BUFF (both hand-labeled).

Training set	Test set	PointNet	PointNet++	DiffusionNet
full dataset	full dataset	85.19	89.13	**91.75**
full dataset	SIZER real scans	74.57	76.89	**81.24**
full dataset	BUFF real scans	77.77	80.68	**83.72**
full dataset + skirts	full dataset + skirts	84.58	87.14	**88.56**
full dataset + skirts	SIZER real scans	75.52	77.57	**79.32**
full dataset + skirts	BUFF real scans	76.83	**80.81**	80.12

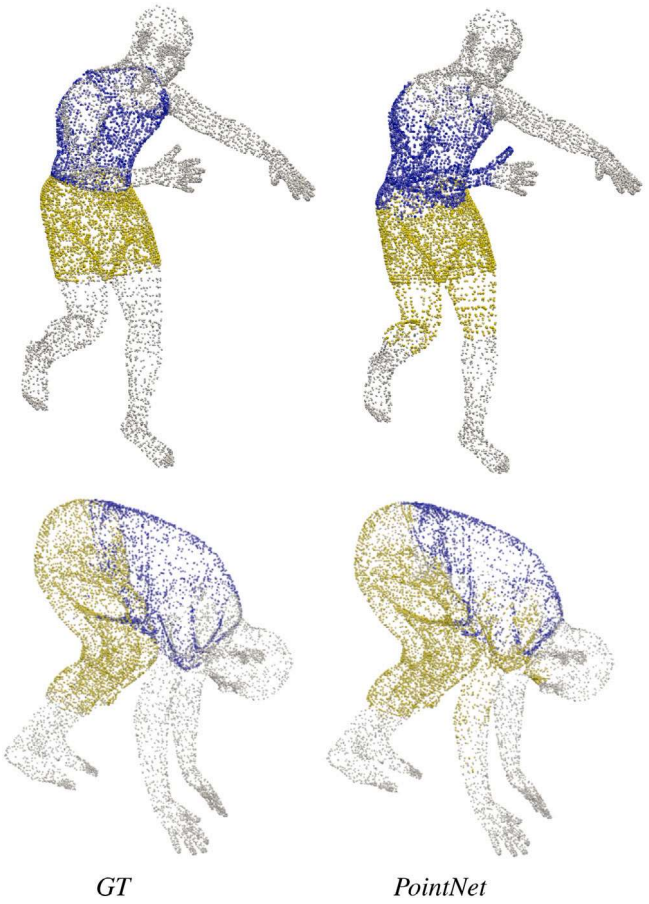

Fig. 8. Some examples of classification failure. The wide variability of poses and the variations of the different garments, even within each class, make the garment segmentation and classification task very challenging.

results very good, around the 80% for the *DiffusionNet*. We will also make available the set of skirt outfits with GIM3D plus.

5. Conclusions

In this work, we have proposed GIM3D plus, i.e., a new synthetic 3D dataset properly designed for dressed human segmentation and garment classification of characters in motion. Our dataset contains over 5000 manifold meshes of many subjects with very different shapes in several poses. The involved samples provide a large variety of accurate details on cloth wrinkles of different garment types. We have shown promising results evaluating our dataset by training state-of-the-art deep learning architectures for dressed human segmentation and garment classification tasks with GIM3D plus. We created our synthetic dataset aiming to emulate the behavior of a real 3D body scan which is a typical scenario where the semantic layers of the human body and the clothes are lost and for which the segmentation task becomes crucial to recover them, while the classification of the garments gives semantic information for the analysis of these type of data. Indeed, it is interesting to note that convincing results have also been shown on the additional test set where real scans have been correctly segmented and classified starting from the training on our synthetic dataset. We believe that GIM3D plus will greatly impact the community since very little data is publicly available for this challenging task.

Limitations In Fig. 8 we can see two examples of classification failure. The wide variety of possible body poses, shapes, and garments make the segmentation and classification tasks very challenging. In particular, all the possible poses of a human body are significant, while the more frequent are way fewer. For example, in the first row of the image, we can see that the failure classified points are in the area of interest of the different pants and sleeves lengths. Even with specific labels on the different types of garments, each class has borderline cases, such as pants just under the knees and sleeves near the elbows. In the second row, we can see a human bent over, we can easily infer that these learning-based methods strongly depend on the training data, and the most frequent motions expect a standing body.

Future works Our GIM3D plus dataset can be easily enriched for matching tasks. As mentioned in Section 4, although the meshes in our dataset are not in point-to-point correspondence, we observe that a unified template (i.e., SMPL) is available for the body under the clothes. In future work, we will exploit this information to define a common template for all 3D models of our dataset. In this way, we should also be able to use data-driven methods to solve the matching problem, another challenge for clothed 3D models.

CRediT authorship contribution statement

Pietro Musoni: Conceptualization, Methodology, Data curation, Writing – original draft. **Simone Melzi:** Conceptualization, Methodology, Writing – review & editing. **Umberto Castellani:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Simone Melzi reports equipment, drugs, or supplies was provided by NVIDIA Corp.

Data availability

Data will be made available on request.

Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the two RTX A5000 GPUs granted to the project titled "Learned representations for implicit binary operations on real-world 2D-3D data", through the Academic Hardware Grant Program to the University of Milano-Bicocca. This work is partially supported by the project of the Italian Ministry of Education, Universities and Research (MIUR) "Dipartimento di Eccellenza 2018–2022" of the Department of Computer Science of Verona University.

References

[1] D. Baraff, A. Witkin, Large steps in cloth simulation, in: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, 1998, pp. 43–54.

[2] R. Bridson, R. Fedkiw, J. Anderson, Robust treatment of collisions, contact and friction for cloth animation, in: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, 2002, pp. 594–603.

[3] R. Narain, A. Samii, J.F. O'brien, Adaptive anisotropic remeshing for cloth simulation, ACM Trans. Graph. 31 (6) (2012) 1–10.

[4] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.

[5] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017).

[6] N. Sharp, S. Attaiki, K. Crane, M. Ovsjanikov, Diffusionnet: Discretization agnostic learning on surfaces, ACM Trans. Graph. 41 (3) (2022) 1–16.

[7] H. Bertiche, M. Madadi, S. Escalera, CLOTH3D: clothed 3d humans, in: European Conference on Computer Vision, Springer, 2020, pp. 344–359.

[8] G. Tiwari, B.L. Bhatnagar, T. Tung, G. Pons-Moll, SIZER: A dataset and model for parsing 3D clothing and learning size sensitive 3D clothing, in: European Conference on Computer Vision, ECCV, Springer, 2020.

[9] C. Zhang, S. Pujades, M.J. Black, G. Pons-Moll, Detailed, accurate, human shape estimation from clothed 3D scan sequences, in: The IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017.

[10] Z. Heming, C. Yu, J. Hang, C. Weikai, D. Dong, W. Zhangye, C. Shuguang, H. Xiaoguang, Deep Fashion3D: A dataset and benchmark for 3D garment reconstruction from single images, in: Computer Vision, ECCV 2020, Springer International Publishing, 2020, pp. 512–530.

[11] P. Musoni, S. Melzi, U. Castellani, GIM3D: A 3D Dataset for Garment Segmentation, in: D. Cabiddu, T. Schneider, D. Allegra, C.E. Catalano, G. Cherchi, R. Scateni (Eds.), Smart Tools and Applications in Graphics - Eurographics Italian Chapter Conference, The Eurographics Association, 2022, http://dx.doi.org/10.2312/stag.20221252.

[12] G. Pons-Moll, S. Pujades, S. Hu, M. Black, ClothCap: Seamless 4D clothing capture and retargeting, :ACM Trans. Graph. (Proc. SIGGRAPH) 36 (4) (2017) http://dx.doi.org/10.1145/3072959.3073711, Two first authors contributed equally.

[13] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M.J. Black, SMPL: A skinned multi-person linear model, ACM Trans. Graph. 34 (6) (2015) 248:1–248:16, http://dx.doi.org/10.1145/2816795.2818013.

[14] S. Wenninger, J. Achenbach, A. Bartl, M.E. Latoschik, M. Botsch, Realistic virtual humans from smartphone videos, in: Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, 2020, pp. 1–11.

[15] lumaAI, 2022, URL https://lumalabs.ai/.

[16] M. Attene, A lightweight approach to repairing digitized polygon meshes, Vis. Comput. 26 (11) (2010) 1393–1406.

[17] J. Huang, Y. Zhou, L. Guibas, ManifoldPlus: A robust and scalable watertight manifold surface generation method for triangle soups, 2020, arXiv preprint arXiv:2005.11621.