STAG: Smart Tools and Applications in Graphics (2022)
D. Cabiddu, T. Schneider, G. Cherchi, and R. Scateni (Editors)

GIM3D: A 3D dataset for garment segmentation

Pietro Musoni[1], Simone Melzi[2], Umberto Castellani[1],

[1]University of Verona [2] University of Milano-Bicocca

Figure 1: *Here some examples of the clothes and the subjects composing the dataset. The 3D meshes of GIM3D 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.*

Abstract

The 3D cloth segmentation task is particularly challenging due to the extreme variation of shapes, even among the same category of clothes. Several data-driven methods try to cope with this problem but they have to face the lack of available data capable to generalize to the variety of real-world data. For this reason, we present GIM3D (Garments In Motion 3D), a synthetic dataset of clothed 3D human characters in different poses. The over 4000 3D models in this dataset are produced by a physical simulation of clothes with different fabrics, sizes, and tightness, using animated human avatars having a large variety of shapes. Our dataset is composed of 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 as a training set on garment segmentation tasks using state-of-the-art data-driven methods for both meshes and point clouds.

CCS Concepts

• ***Computing methodologies*** → ***Shape analysis;*** • ***Theory of computation*** → *Computational geometry;* • ***Mathematics of computing*** → *Functional analysis;*

DOI: 10.2312/stag.20221252

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 [BW98, BFA02, NSO12] 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[†]). 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 the 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 to be used for modeling and animation purposes. 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. This task is 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 in 3D segmentation consists of adopting learning-based technologies that exploit data-driven paradigms. In particular, new neural network architectures has been proposed for the segmentation on the 3D domain represented by simple 3D point clouds [QSMG17, QYSG17] or triangular meshes [SACO22]. To extend this learning methodology to dressed humans a crucial need is the availability of an appropriate dataset of labeled samples.

In this paper, we proposed GIM3D which is a novel dataset of dressed-human for cloth segmentation. Our GIM3D is composed of several human subjects with different geometric characteristics. Each subject is wearing different garment types, i.e, t-shirts, tops, trousers 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 dataset has been created from a selection of digital garments available on the CLOTH3D [BME20]. A well-designed processing 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 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 [TBTPM20, ZPBPM17, BME20, HYH*20] but they have been not specifically designed for cloth segmentation. For instance, in [ZPBPM17] the BUFF dataset has been proposed to estimate the human shape under clothes and therefore cloth labeling is not available. In [BME20, HYH*20] the CLOTH3D and DeepFashion datasets provide the garments in separated meshes for the estimation of 3D cloth from a single image. In [TBTPM20] the SIZER dataset has been introduced to learn the generation of synthetic garments of different sizes.

Our main contribution is threefold:

- we provide the first dataset of labeled dressed humans that has been specifically designed for cloth segmentation. The dataset is publicly available in[‡].
- we show that our dataset is largely expressive to enable a neural network to reliably learn the cloth segmentation task. Three state-of-the-art neural network architectures for 3D segmentation have been evaluated, namely PointNet [QSMG17], PointNet++ [QYSG17], and DiffusionNet [SACO22].
- 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 the state of the art on publicly available datasets of clothed 3D. In Section 3 we introduce our GIM3D 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.

2. Related Works

The main motivation for our work is the lack of available data for 3D garment segmentation, which is a real challenge for the computer graphics community. Some really interesting datasets are shared by the community built for different tasks, such as human 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 [TBTPM20] 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* [PMPHB17], a parametric 3D model of clothed characters. The aim of the work is to predict the clothing over an avatar

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

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

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) [ZPBPM17] 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 seconds (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 [LMR*15] 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 tasks.

DeepFashion 3D Deep Fashion3D [HYH*20] 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 [BME20] 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, varying the shape parameters that allows to create a large variety of human models. 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.

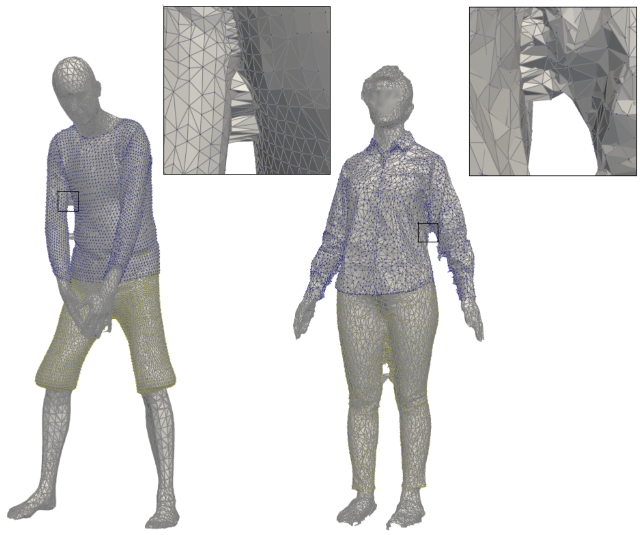


Figure 2: *Here is an example of artifacts on the dataset shapes. On the left, we have a synthetic model from our dataset and on the right a real 3D scan from the SIZER dataset, we highlight how the process to obtain a single manifold mesh produces artifacts similar to the output meshes of 3D scanners.*

All these works have peculiarities very useful for the different tasks in which they are involved. So we decided to build our own dataset for our segmentation target, we use the huge variety of clothes and the realistic dynamics of the garments in CLOTH3D to build our dataset GIM3D. 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.

3. 3D Garments In Motion (GIM3D) 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[§] 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. 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 mesh the avatar and the separate garments in order to create a synthetic scan.

§ www.mocap.cs.cmu.edu/

Figure 3: *Some examples of the three subsets of our dataset. At the top some samples of the top + trouser dataset, in the middle the top + trouser 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 t-shirt label we can see several lengths of the sleeves, in the top category in some cases there are shoulder straps in some cases none and the lengths of the trousers of each category varies from over the knees to the ankles.*

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* [Att10]. This step helps the following step to create a watertight closed mesh.
- The method presented in [HZG20] 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 figure 4) (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 figure 4

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 (figure 2). 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 pre-processing 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.

Data Description. The dataset is composed of a total of 4623 meshes, divided into 1851 with two separate outfits (shirts and trousers) and 2772 with a single garment (jumpsuit). We decided to use three typologies of outfits: i) t-shirts and trousers, ii) top and trousers, and iii) jumpsuit. In figure 3 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, depending also 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 figure 3. 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 dataset, we took a single random pose available from the motion sequences. We will share publicly the code that implements our procedure to create the output meshes from the source

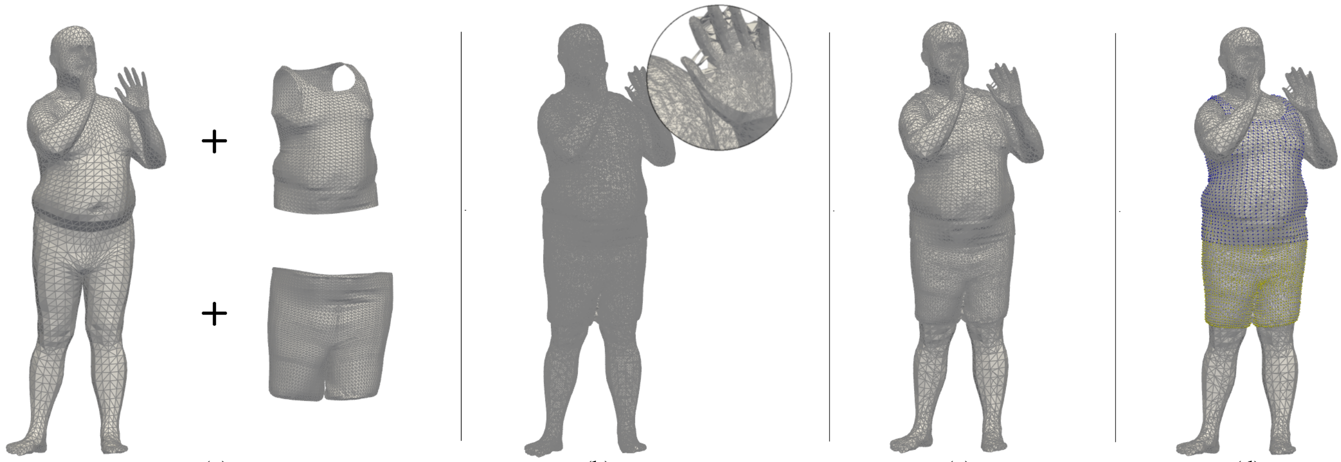


Figure 4: *Here the pipeline to produce our dataset. In (a) the 3 separate 3D models coming from CLOTH3D dataset, then (b) we merge the body with the clothes and we obtain a very dense watertight mesh. In (c) lower the number of vertices maintaining the details. In the last step (d) we use the separate meshes of the clothes to apply the labels on the mesh using nearest neighbor algorithm.*

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

Table 1: *Here the accuracy 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, in the second column the sets of data on which we test the methods. For the training set we use* t-shirt + trouser *which is composed by sleeved shirts,* top + trouser *has only tanktops and tops without shoulder straps for the upper clothes. The* complete dataset *is a merge of the first two datasets and the* binary labels dataset *is composed by 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.*

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 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.

4. Experiments

We tested our GIM3D dataset on state-of-the-art methods for 3D segmentation. We evaluated the results using both the entire dataset and its different subsets. As surface segmentation methods we assessed the three following deep learning architectures: i) *PointNet* [QSMG17], ii) *PointNet++* [QYSG17] and iii) *DiffusionNet* [SACO22]. The two first networks operate on Point Clouds meanwhile the third one takes as input a polygonal mesh. As shown in table 1 we followed four different training scenarios, we train the three different networks using:

- the *t-shirt + trouser dataset*, composed of 815 meshes in t-shirt and trouser outfit,

Training set	PointNet	PointNet++	DiffusionNet
t-shirt + trouser	85.40	88.21	87.17
top + trouser	73.93	75.69	75.01

Table 2: *Here the accuracy results on clothes and body segmentation with PointNet, PointNet++ and DiffusionNet architectures. In this experiment we use the two subset for training set and we crossed the categories for the test set (testing t-shirts on the networks trained with tops and vice versa).*

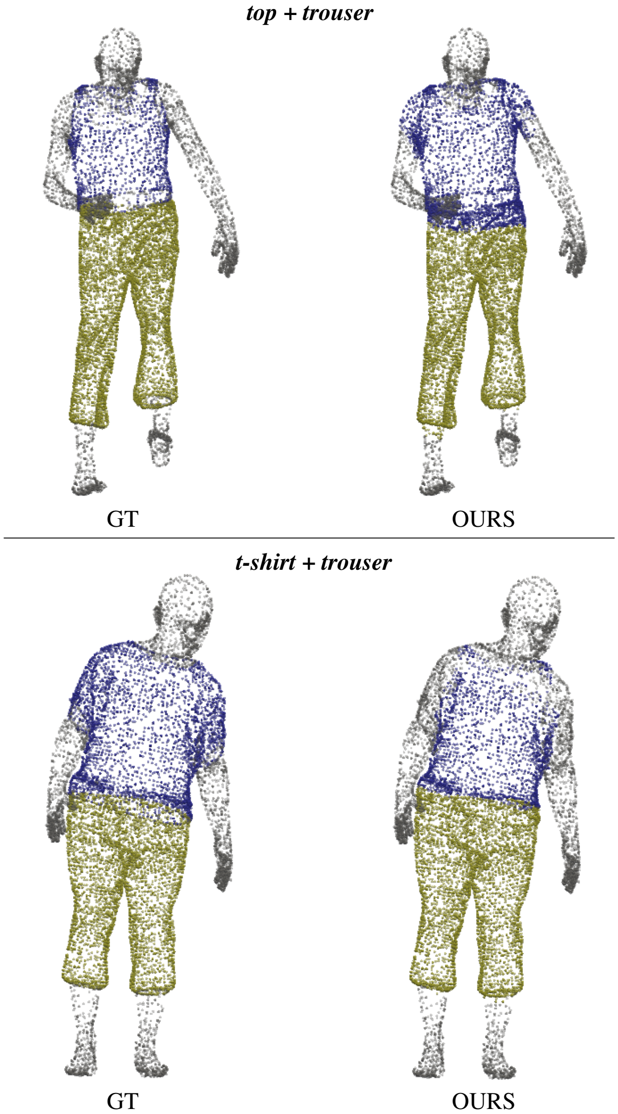


Figure 5: *Here is an example of the segmentation experiment reported in table 2 where the left column is the GT and right column is the result of the test. In the first row, we have an example of segmentation where the training set is the* t-shirt + trouser *dataset and the test set is the* top + trouser *dataset, in the second row we have the opposite experiment. These examples are both results of* PointNet++ *method, which provides the best performance.*

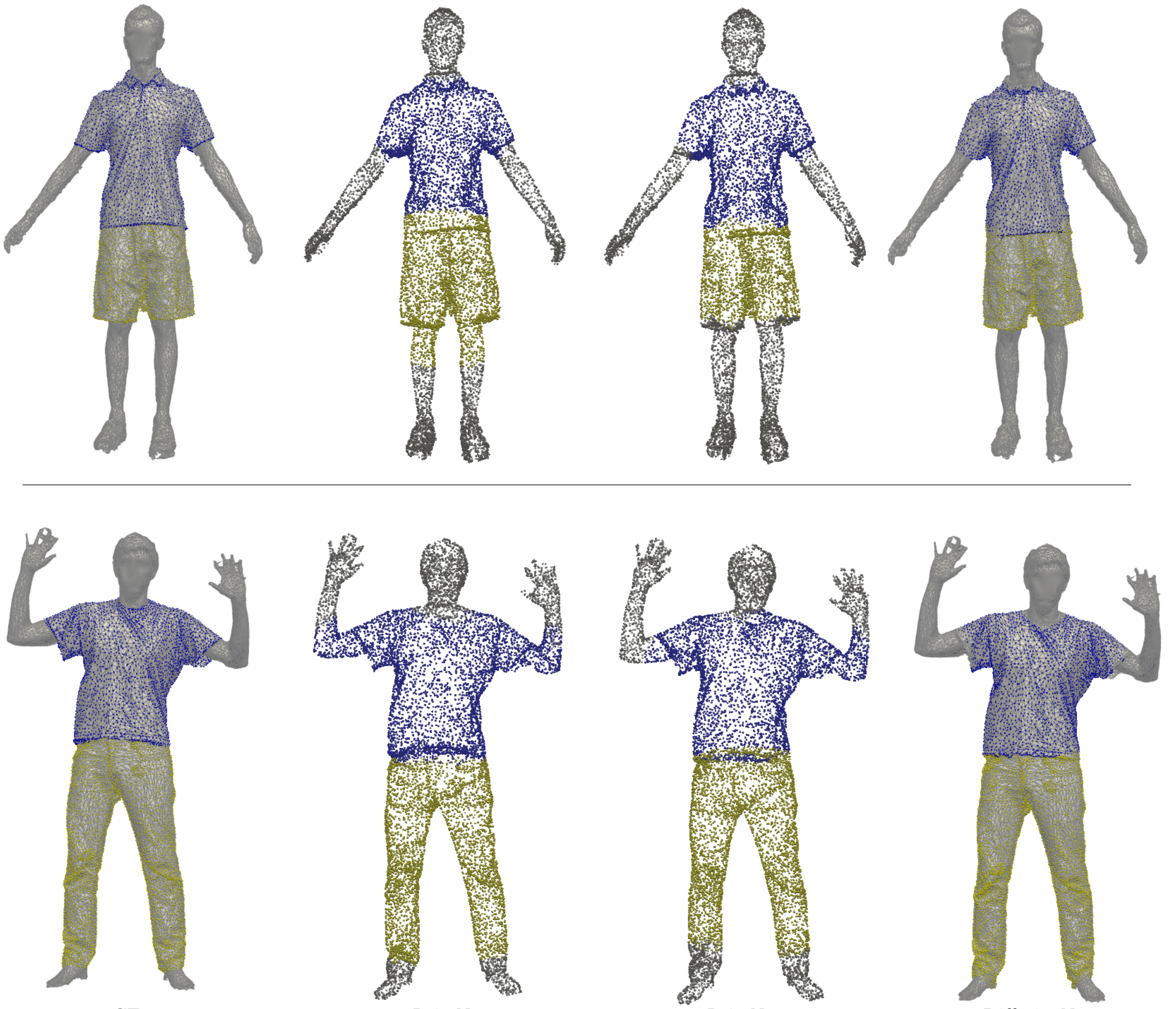
- the *top + trouser dataset*, which has 1036 subjects with singlets or tops with trousers,
- the *complete dataset*, that is the fusion of the two previous datasets (with 1851 3D shapes), and
- the *jumpsuit dataset*, which is composed of only one garment for the whole body and can be seen as the fusion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

We test the three methods with different set of data, in the first part of the 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 set of experiments we test the three methods on shapes from SIZER and BUFF datasets. In table 2 we make a cross experiment between the two different outfits, so we test the *top + trouser* data on the architectures trained with *t-shirt + trouser* 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 t-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 figure 5. We use 80/20 train-test split for all the experiments (e.g. for the complete dataset the training and the test sets are composed of 1480 and 371 shapes respectively). In figure 3 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 varies in the length of trousers and shirts. The labels for this small test-set have been manually defined.

In figure 6 some qualitative results from these experiments and in table 1 the accuracy of the networks on these two test sets is shown. The *complete dataset* has inferior results in relation to the subsets which is composed of, 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 *t-shirt + trouser*, 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 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 labeled generally 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 figure 6, 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* perform better with meshes (despite can take also pointcloud as input). *PoinNet++* uses also the information of the points normals, respect to the *PoinNet* architecture, and in the case of meshes they can easily estimated. For these reasons these methods performs better then the basic *PoinNet* method. As exposed in table 1.

5. Conclusions

In this work, we have proposed GIM3D, i.e., a new synthetic 3D dataset properly designed for garment segmentation of dressed humans in motion. Our dataset contains over 4000 manifold meshes of a large number of subjects with very different shapes in several



GT *PointNet* *PointNet++* *DiffusionNet*

Figure 6: *Here two examples of real scans from BUFF and SIZER datasets labeled using the three networks used for the experiments, and our dataset as training set. In the first row a SIZER scan (every scan in SIZER is in A-pose) and in the second row the scan from BUFF in 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 operates on Point clouds,* DiffusionNet*, meanwhile, works with meshes.*

different 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 the 3D segmentation task with GIM3D. 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. Indeed, it is interesting to note that convincing results have been shown also on the additional test-set where real scans have been correctly segmented starting from the training on our synthetic dataset. We believe that GIM3D will give a great impact on the community since there is very little data publicly available for this very challenging task.

Future work

Our GIM3D 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. As future work we propose to work to exploit this information to define a common template for all 3D models of our dataset. In this way we should be able to use data-driven methods also to solve the matching problem, which is another challenge for clothed 3D models.

References

[Att10] ATTENE M.: A lightweight approach to repairing digitized polygon meshes. *The visual computer 26*, 11 (2010), 1393–1406.

[BFA02] BRIDSON R., FEDKIW R., ANDERSON J.: 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.

[BME20] BERTICHE H., MADADI M., ESCALERA S.: Cloth3d: clothed 3d humans. In *European Conference on Computer Vision* (2020), Springer, pp. 344–359.

[BW98] BARAFF D., WITKIN A.: Large steps in cloth simulation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), pp. 43–54.

[HYH*20] HEMING Z., YU C., HANG J., WEIKAI C., DONG D., ZHANGYE W., SHUGUANG C., XIAOGUANG H.: Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In *Computer Vision – ECCV 2020* (2020), Springer International Publishing, pp. 512–530.

[HZG20] HUANG J., ZHOU Y., GUIBAS L.: Manifoldplus: A robust and scalable watertight manifold surface generation method for triangle soups. *arXiv preprint arXiv:2005.11621* (2020).

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graph. 34*, 6 (2015), 248:1–248:16. doi:10.1145/2816795.2818013.

[NSO12] NARAIN R., SAMII A., O'BRIEN J. F.: Adaptive anisotropic remeshing for cloth simulation. *ACM transactions on graphics (TOG) 31*, 6 (2012), 1–10.

[PMPHB17] PONS-MOLL G., PUJADES S., HU S., BLACK M.: Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics, (Proc. SIGGRAPH) 36*, 4 (2017). Two first authors contributed equally. URL: http://dx.doi.org/10.1145/3072959.3073711.

[QSMG17] QI C. R., SU H., MO K., GUIBAS L. J.: 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.

[QYSG17] QI C. R., YI L., SU H., GUIBAS L. J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems 30* (2017).

[SACO22] SHARP N., ATTAIKI S., CRANE K., OVSJANIKOV M.: Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG) 41*, 3 (2022), 1–16.

[TBTPM20] TIWARI G., BHATNAGAR B. L., TUNG T., PONS-MOLL G.: Sizer: A dataset and model for parsing 3d clothing and learning size sensitive 3d clothing. In *European Conference on Computer Vision (ECCV)* (August 2020), Springer.

[ZPBPM17] ZHANG C., PUJADES S., BLACK M. J., PONS-MOLL G.: Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017).