End-to-end Learning of Body Weight Prediction from Point Clouds with Basis Point Sets

Alexander Bigalke[1], Lasse Hansen[1], Mattias P. Heinrich[1]

[1]Institut für Medizinische Informatik, Universität zu Lübeck
alexander.bigalke@uni-luebeck.de

Abstract. The body weight of a patient is an important parameter in many clinical settings, e.g. when it comes to drug dosing or anesthesia. However, assessing the weight through direct interaction with the patient (anamnesis, weighing) is often infeasible. Therefore, there is a need for the weight to be estimated in a contactless way from visual inputs. This work addresses weight prediction of patients lying in bed from 3D point cloud data by means of deep learning techniques. Contrary to prior work in this field, we propose to learn the task in an end-to-end fashion without relying on hand-crafted features. For this purpose, we adopt the concept of basis point sets to encode the input point cloud into a low-dimensional feature vector. This vector is passed to a neural network, which is trained for weight regression. As the originally proposed construction of the basis point set is not ideal for our problem, we develop a novel sampling scheme, which exploits prior knowledge about the distribution of input points. We evaluate our approach on a lying pose dataset (SLP) and achieve weight estimates with a mean absolute error of 4.2 kg and a mean relative error of 6.4 % compared to 4.8 kg and 7.0 % obtained with a basic PointNet.

1 Introduction

The precise knowledge of a patient's body weight is a crucial requirement in several clinical scenarios, including anesthesia or drug dosage. In emergency situations, however, patients are often unable to communicate their weight due to unconsciousness, dementia or neurological disorder. Weighing the patient on-site with an ordinary scale is infeasible in case of severe injuries, and bed scales are expensive and not always available. For these reasons, weight is often estimated by clinical staff although this procedure has been shown to be error-prone in clinical studies [1].

To obtain more accurate weight estimates, several works use a multiple linear regression model to infer body weight from biometric measurements such as height, and waist and hip circumference [2]. Since manual measurements of these quantities are time-intensive and infeasible in case of certain injuries, it is difficult to integrate this approach into clinical routine. Instead, a fully automatic and contactless weight estimate is desirable.

This can be achieved by deriving the weight estimate from visual sensor data using methods from computer vision. For this purpose, the use of depth sensors is particularly suitable. Firstly, depth maps and corresponding point clouds carry rich geometric information which is of eminent importance for accurate weight estimates. Secondly, patients are unidentifiable on depth maps which prevents any privacy concerns.

This work addresses the task of weight estimation of patients lying in bed from 3D point cloud data by means of deep learning techniques. Contrary to prior work in this field, we aim to learn weight prediction in an end-to-end fashion.

1.1 Related work

More generic work in the field of weight estimation from visual data predicts the weight of free-standing subjects from RGB-D data [3]. The proposed method segments the subject from the background and extrapolates biometric measures from the silhouette. The deduced measures are fed into a neural network to regress the subject's weight.

Several works address weight estimation from point clouds of lying patients in a clinical environment [4,5]. Libra3D [4] fits a mesh to the point cloud whereby the patient's back is modeled with help of the bed plane. Based on the mesh, the volume of the patient is calculated and multiplied with a fixed empirically determined density to obtain a weight estimate. The authors of [5] extend this work by additionally extracting more abstract features from the point cloud, which are forwarded by a neural network for weight regression. All of these works rely on hand-crafted features and are not trained end-to-end.

In recent years, end-to-end learning from point clouds has become viable owing to deep learning architectures that directly operate on raw point sets. The pioneering PointNet [6] applies a shared multi-layer perceptron to each point individually and achieves permutation invariance through a symmetric max pooling operation.

1.2 Contribution

To our knowledge, this is the first work to learn weight prediction from 3D point clouds in an end-to-end fashion. Since learning weight regression directly from raw point clouds using a PointNet architecture [6] is a complex task, we suggest to simplify the problem by considering input point clouds relative to a fixed reference. To achieve this, we adopt the idea of basis point sets (BPS) [7] to encode the input point cloud. The resulting feature vector is subsequently fed into a fully connected neural network to regress the weight. Based on the observation that the construction of the BPS in [7] is not ideal for our specific problem, we propose an adapted sampling scheme to incorporate prior knowledge about the distribution of input points. We experimentally validate our approach on the SLP dataset [8] and significantly outperform several baselines, including a PointNet architecture [6].

2 Materials and methods

Our method receives a point cloud cropped around the bed as input and outputs the patient's weight in kg. The point cloud is initially pre-processed, subsequently encoded by means of a BPS and finally processed by a neural network for weight prediction (Fig. 1). We assume the patient to be uncovered and in a supine position, which can easily be realized in clinical workflow.

2.1 Pre-processing

In the pre-processing step, the patient needs to be segmented from the bed. First, we use the RANSAC algorithm [9] to fit a plane to the mattress and keep only the points above the plane as it was done in [4]. Most of the kept points belong to the patient, but there may remain point clusters belonging to other objects on the bed. To remove those points, we cluster the cloud using DBSCAN [10] and only keep the largest cluster.

2.2 Basis point set and neural network

After pre-processing, we are left with a set of patient point clouds $\boldsymbol{X}_i \in \mathbb{R}^{N_i \times 3}$, $(i = 1, ..., p)$, each comprising N_i points $\boldsymbol{x}_{ij} \in \mathbb{R}^3$. We encode the clouds with help of a BPS as elaborated in [7]. In [7], each cloud is initially normalized to fit a unit sphere which entails a loss of scale information. Since scale is indispensable for weight estimation, we only mean-center each cloud. Subsequently, a BPS

$$\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_k], \boldsymbol{b}_j \in \mathbb{R}^3, \|\boldsymbol{b}_j\| \leq r \tag{1}$$

is constructed by uniform sampling of k points from a sphere of radius r. This set is fixed for all point clouds in training and test set. We select $k = 2048$ and set the radius to the maximal radius of all point clouds in the training set, i.e. $r = \max_i(\max_j \|\boldsymbol{x}_{ij} - (\sum_k \boldsymbol{x}_{ik})/N_i\|)$.

Given the BPS $\boldsymbol{B}$, an input point cloud $\boldsymbol{X}_i$ is encoded by computing the distance from each basis point to the nearest point in the input cloud, yielding

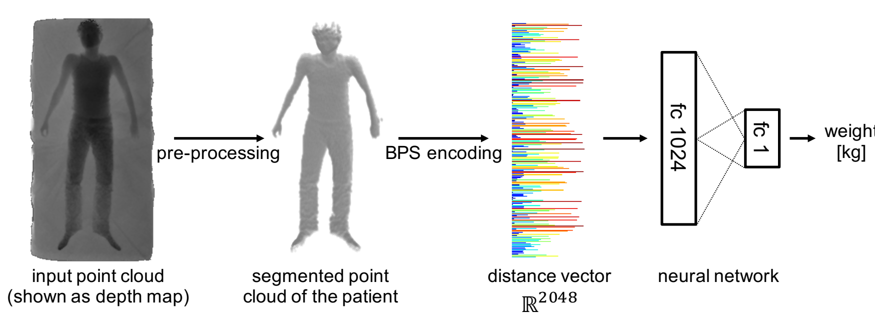


Fig. 1. Overview of our proposed pipeline for weight estimation from point clouds.

a k-dimensional feature vector

$$\boldsymbol{f}_i^{\boldsymbol{B}} = [\min_{\boldsymbol{x}_{ij} \in \boldsymbol{X}_i} d(\boldsymbol{b}_1, \boldsymbol{x}_{ij}), ..., \min_{\boldsymbol{x}_{ij} \in \boldsymbol{X}_i} d(\boldsymbol{b}_k, \boldsymbol{x}_{ij})] \in \mathbb{R}^k \tag{2}$$

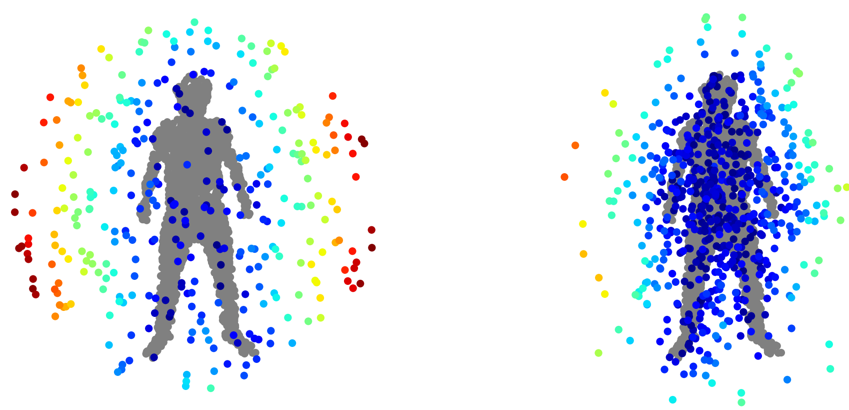
This feature vector is subsequently fed into a neural network, consisting of the following sequence of layers: BN, FC(1024), ReLU, BN, Dropout(p=0.8), FC(1). The network parameters are optimized by minimizing a mean squared error loss between predicted weight and ground truth.

2.3 Adapted sampling of basis points

In Fig. 2(a), a BPS obtained by uniform sampling in the sphere is shown relative to an input point cloud. We observe that many basis points are far away from the patient and thus encode less detailed information. As all patients have a similar orientation and occupy similar regions of the sphere, we conclude that the uniform distribution of basis points is not ideal for our specific problem. We believe that a more expressive basis can be constructed by incorporating prior knowledge about the distribution of input points. To achieve this, we propose to sample the basis points from a unified point cloud which comprises all clouds from the training set. As this basis is prone to overfitting, we subsequently add Gaussian noise with a standard deviation of $\sigma = 0.3$ to the sampled basis points. The resulting BPS is depicted in Fig. 2(b).

3 Results

Dataset. We evaluate our method on a subset of the SLP dataset [8]. The subset comprises depth maps of 109 subjects which are lying in bed in a supine position



(a) Uniform sampling in the sphere. (b) Sampling from training points.

Fig. 2. Comparison of two basis point sets constructed with different sampling schemes. We visualize a slice of the sphere around the input point cloud of a patient, which is shown in gray for reference. Basis points are shown in colour to represent the distance to the closest input point. The basis points constructed by our sampling scheme (b) are substantially more concentrated around the patient.

Table 1. Results for weight estimation on the SLP dataset.

method	MAE [kg]	MRE [%]	in 10 % range [%]
mean	9.46	14.6	40.8
median	9.54	14.3	44.9
PointNet [6]	5.42 ± 0.4	8.0 ± 0.5	70.7 ± 3.2
PointNet [6] & median	4.84 ± 0.48	7.1 ± 0.6	74.7 ± 5.3
BPS random sampling	4.91 ± 0.09	7.5 ± 0.1	74.0 ± 0.8
BPS adapted sampling	4.69 ± 0.08	7.1 ± 0.1	76.1 ± 1.0
BPS adapted sampling & median	$\mathbf{4.19 \pm 0.12}$	$\mathbf{6.4 \pm 0.2}$	$\mathbf{78.6 \pm 2.9}$

without a cover. Each subject takes 15 different poses while staying in supine position, yielding an overall of 1635 frames. For each frame, a bounding box around the bed is obtained with the help of depth thresholding, and the corresponding image crop is transformed to a point cloud using the internal camera parameters. The weight of the subjects ranges from 43.7 to 105.1 kg with a mean of 68.0 kg and a standard deviation of 12.7 kg. We use the first 60 subjects for training and results are reported for the remaining 49 subjects.

Implementation Details. For pre-processing, we run RANSAC with a threshold of 1 cm for 1000 iterations. DBSCAN is used with $\epsilon = 2.5$ cm and minpts = 5. Network parameters are optimized with the ADAM optimizer. The initial learning rate is set to 0.001 and halved every 40 epochs. We use a batch size of 16 and train for 200 epochs. Each experiment is repeated ten times and we report mean and standard deviation.

Baselines. As baseline, we train a basic PointNet [6] to directly regress the weight from the point cloud of the patient. Additionally, we estimate the weight of each test subject with a constant value which corresponds to the mean/median weight of all subjects from the training set of the same sex as the test subject.

Results are presented in Tab. 1. We compare the baseline methods to three variants of our approach: 1) uniform sampling of basis points in the sphere, 2) sampling the basis points from training points, 3) same as 2), but for each subject, we take the median of the predicted weights for all 15 frames. For each method, we report the following metrics on the test set: mean absolute error (MAE), mean relative error (MRE), percentage of subjects within a relative error range of ± 10 %.

Results demonstrate that BPS with random sampling halves both MAE and MRE of the mean/median baselines and considerably improves on the PointNet architecture without median filtering as well. Applying our adapted sampling further reduces MAE by 0.22 kg and MRE by 0.4 % points. Finally taking the median of 15 independent weight estimates for the same subject yields another improvement of 0.5 kg in MAE and 0.7 % points in MRE. That way, we achieve an overall MAE of 4.19 kg, MRE of 6.4 % and the weight of 78.6 % of the subjects is estimated within a 10 % error range. This constitutes a relative performance gain in MAE of 56 % compared to the mean/median baseline and of 13.4 % in relation to the corresponding PointNet model.

4 Discussion

This work successfully applied the concept of BPS [7] to learn body weight prediction from point clouds of lying patients in an end-to-end fashion. We optimized the method for the specific problem at hand by introducing a customized sampling scheme for basis construction which takes the prior distribution of input points into account and thus contributed a meaningful performance gain. Finally, the experiments showed that a further increase of accuracy can be achieved by statistical averaging over several independent weight estimates for the same subject. Altogether, our method achieves a higher accuracy (MAE=4.2 kg, MRE=6.4 %) than weight estimates by clinical staff, which exhibit MAEs between 5.7 and 8.7 kg in [2] and MREs of 7.7 to 11.0 % in [1]. That way, our work demonstrates the potential of end-to-end deep learning in the context of weight estimation and thus encourages further research in this direction. Future work could, for instance, incorporate semantic labels or point descriptors into the encoding or address the construction of an even more tailored basis set.

Acknowledgement. We gratefully acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy of Germany (FKZ: 01MK20012B).

References

1. Menon S, Kelly AM. How accurate is weight estimation in the emergency department? Emerg Med Australas. 2005;17(2):113–116.
2. Lorenz MW, Graf M, Henke C, et al. Anthropometric approximation of body weight in unresponsive stroke patients. J Neurol Neurosurg Psychiatry. 2007;78(12):1331–1336.
3. Velardo C, Dugelay JL. What can computer vision tell you about your weight? Procs IEEE EUSIPCO. 2012; p. 1980–1984.
4. Pfitzner C, May S, Merkl C, et al. Libra3D: Body weight estimation for emergency patients in clinical environments with a 3D structured light sensor. IEEE Int Conf Robot Autom. 2015; p. 2888–2893.
5. Pfitzner C, May S, Nüchter A. Neural network-based visual body weight estimation for drug dosage finding. Medical Imaging 2016: Image Processing. 2016;9784:97841Z.
6. Qi CR, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2017; p. 652–660.
7. Prokudin S, Lassner C, Romero J. Efficient learning on point clouds with basis point sets. Proc IEEE Int Conf Comput Vis Workshop. 2019; p. 0–0.
8. Liu S, Ostadabbas S. Seeing under the cover: A physics guided learning approach for in-bed pose estimation. Proc Springer Int Conf Med Image Comput Comput Assist Interv. 2019; p. 236–245.
9. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 1981;24(6):381–395.
10. Ester M, Kriegel HP, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd. 1996;96(34):226–231.