



Accurate and Robust Patient Height and Weight Estimation in Clinical Imaging Using a Depth Camera

Birgi Tamersoy¹(✉), Felix Alexandru Pirvan², Santosh Pai³,
and Ankur Kapoor³

¹ Digital Technology and Innovation, Siemens Healthineers, Erlangen, Germany
`birgi.tamersoy@siemens-healthineers.com`

² Siemens S.R.L. Romania, Bucuresti, Romania
`felix.pirvan@siemens.com`

³ Digital Technology and Innovation, Siemens Healthineers, Princeton, NJ, USA
`{santosh.pai,ankur.kapoor}@siemens-healthineers.com`

Abstract. Accurate and robust estimation of the patient's height and weight is essential for many clinical imaging workflows. Patient's safety, as well as a number of scan optimizations, rely on this information. In this paper we present a deep-learning based method for estimating the patient's height and weight in unrestricted clinical environments using depth images from a 3-dimensional camera. We train and validate our method on a very large dataset of more than 1850 volunteers and/or patients captured in more than 7500 clinical workflows. Our method achieves a PH5 of 98.4% and a PH15 of 99.9% for height estimation, and a PW10 of 95.6% and a PW20 of 99.8% for weight estimation, making the proposed method state-of-the-art in clinical setting.

Keywords: Height Estimation · Weight Estimation · Clinical Workflow Optimization

1 Introduction

Many clinical imaging workflows require the patient's height and weight to be estimated in the beginning of the workflow. This information is essential for patient's safety and scan optimizations across modalities and workflows. It is used for accurate prediction of the Specific Absorption Rate (SAR) in Magnetic Resonance Imaging (MRI), contrast dose calculations in Computed Tomography (CT), and drug dose computations in Emergency Room (ER) workflows.

Contrary to its importance, there are no widely established methods for estimating the patient's height and weight. Measuring these values using an actual scale is not a common clinical practice since: 1) a measurement scale is not available in every scan room, 2) manual measurements add an overhead to the clinical workflow, and 3) manual measurements may not be feasibly for some

patients with limited mobility. Alternative methods such as the Lorenz formulae [1] or the Crandall formulae [2] need additional body measurements (e.g. mid-arm circumference, waist circumference and/or hip circumference) and are neither very accurate nor simple. Consequently, clinical staff usually relies either on previously recorded patient information or their own experience in estimating the patient’s height and weight, where the estimated values may significantly deviate from the actual values in both cases.

In this paper we present a deep-learning based method for accurately and robustly estimating the patient’s height and weight in challenging *and* unrestricted clinical environments using depth images from a 3-dimensional (3D) camera. We aim to cover the patient demographics in common diagnostic imaging workflows. Our method is trained and validated on a very large dataset of more than 1850 volunteers and/or patients, captured in more than 7500 clinical scenarios, and consists of nearly 170k depth images. We achieve a PH5 (percentage of the height estimates within 5% error) of 98.4% and a PH15 of 99.9% for height estimation, and a PW10 (percentage of the weight estimates withing 10% error) of 95.6% and a PW20 of 99.8% for weight estimation, making the proposed method state-of-the-art in clinical setting.

In addition to the clinical significance, our method has the following primary technical novelties: 1) we formulate the problem as an end-to-end single-value regression problem given only depth images as input (i.e. no error-prone intermediate stages such as volume computations), 2) we present a multi-stage training approach to ensure robustness in training (i.e. no need for hyper-parameter tunings at any stage), and 3) we evaluate our method on a very large dataset of both volunteers and patients, using 23-fold cross validation to ensure field generalization.

2 Related Work

A large number of previous methods have been proposed and independently evaluated for patient height and weight estimation. Since patient height estimation is considered to be an easier problem, the primary focus of the previous work has been on patient weight estimation.

Most of the existing work in patient weight estimation are formulae-based approaches where one or more anthropometric measurements are used for estimating the patient’s weight. The Mercy method uses the humeral length and the mid-arm circumference (MAC) for estimating the paediatric body weight [4]. PAWPER XL-MAC is another height-based (in combination with MAC) method for estimating the body weight in paediatric patients [3]. Broca index [6] and Kokong formula [8] do not take into account a person’s body habitus and estimates the “ideal weight” using only the body height information. Buckley method [7], Lorenz formulae [1], Crandall formulae [2] provide gender-specific weight estimation formulae given some anthropometric measurements such as the abdominal circumference, tight circumference, and MAC.

Formulae-based approaches are independently evaluated in numerous studies, both for paediatric patients [3] and adult patients [9,10]. A common conclusion of these studies is that the formulae-based methods usually perform poorly in the clinical setting with PW10 values below 70%.

More recently several methods have been proposed that leverage 3D camera input for estimating the patient’s weight. In [11] an RGB-D camera and a thermal camera are used for precisely segmenting the patients and then extracting volume based features. These features are then fed into an artificial neural network (ANN) for patient weight estimation. In [12], first a 3D patient avatar is fitted to the acquired depth images, which is then used for part-volume based weight estimation. Both of these approaches require a number of additional algorithmic steps and the challenges of the clinical setup (such as heavy occlusions due to covers and/or additional devices like coils during an MRI examination) may affect the accuracy of the results.

Estimation of the patient weight by the clinical staff remains to be the most common approach in the clinical workflow. In [15], the performance of clinical staff is determined as PW10 of 78% for nurses and PW10 of 59% for physicians. In [16] the performance of the clinical staff is determined as PW10 of 66% for both nurses and physicians.

As a clinical acceptance criteria, Wells et al. [3] proposes a minimum accuracy for patient weight estimation as PW10 greater than 70% and PW20 greater than 95%.

3 Approach

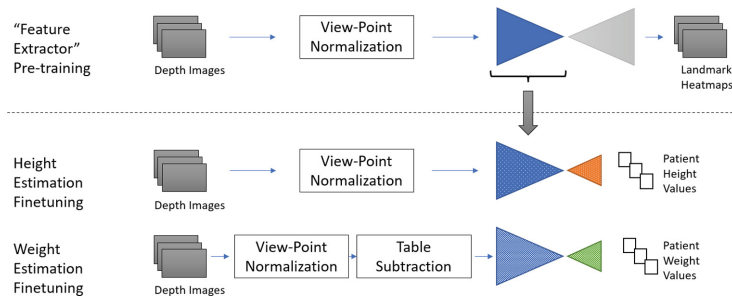


Fig. 1. Overview of the proposed method.

Overview of the proposed method is illustrated in Fig. 1. Our method takes in “normalized” depth images as input. This normalization covers two aspects: 1) normalization with respect to the view-point of the depth camera, and 2) normalization with respect to the variations in the patient tables. Input normalized depth images are then fed into a common feature extraction encoder network.

This encoder network is trained using landmark localization as an auxiliary task. In the last stage we train and utilize two separate single-value regression decoder networks for estimating the patient’s height and weight. These steps are explained in-detail in the following sub-sections.

3.1 Obtaining “Normalized” Depth Images

When training deep neural networks it is more data efficient to eliminate as much of the foreseen variances in the problem as possible in the preprocessing steps. Conceptually this can be thought as reducing the “dimensionality” of the problem before the model training even starts.

Camera view-point is a good example when model inputs are images. With a known system calibration, a “virtual camera” may be placed in a consistent place in the scene (e.g. with respect to the patient table) and the “re-projected” depth images from this virtual camera may be used instead of the original depth images. This way the network training does not need to learn an invariance to the camera view-point, since this variance will be eliminated in the preprocessing. This process forms the first step in our depth image normalization. Figure 2 presents some examples.

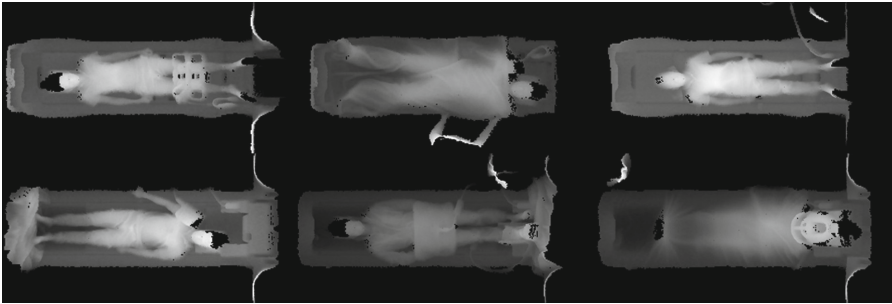


Fig. 2. Example re-projected depth images. View-point normalization simplifies the problem for the deep neural networks.

In the second step, we consider the back-surface of a patient which is not visible to the camera. In a lying down pose, the soft-tissue deforms and the back-surface of the patient takes the form of the table surface. Since there are a variety of patient tables (curved or flat) we eliminate this variance by performing a “table subtraction” from the view-point normalized depth images. This is especially important for accurate patient weight estimation across different systems.

For table subtraction, top surfaces extracted from 3D models of the corresponding patient tables are used. For a given input image, it is assumed that the corresponding patient table is known since this information is readily available in an integrated system. Even though we leveraged the actual 3D models of the patient tables, since the proposed approach only requires the top surface, this

information may also be obtained during calibration by taking a depth image of the empty patient table. Figure 3 presents some examples of the table subtraction process.

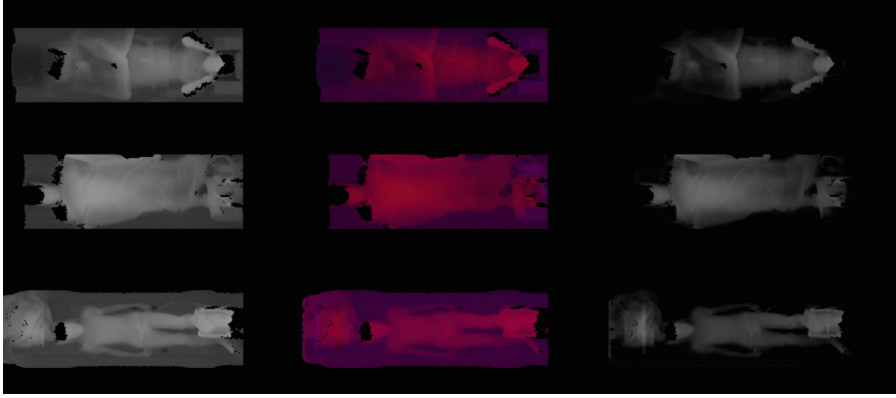


Fig. 3. “Table subtraction” for patient back-surface consistency. Left - original inputs. Center - original inputs overlaid with automatically aligned patient tables. Right - table subtracted depth images.

3.2 Learning Accurate Low-Level Features

Single-value regression problems require more attention during the training since the limited feedback provided through the loss function may result in the collapse of some of the features in the lower levels of the network, especially if a larger model is being trained.

In order to ensure that the learned low-level features are high-quality, we start with the training of a standard image-in image-out encoder-decoder network using the landmark localization as an auxiliary task. With this task, the network is expected to both capture local features (for getting precise landmark locations) *and* holistic features (for getting globally consistent landmark locations) better than the case where the network was to asked only to regress a single-value such as the height or the weight of the patient.

Once the encoder-decoder network is trained, we disregard the decoder part and use the encoder part as the pre-trained feature-extractor.

3.3 Task-Specific Decoders for Height and Weight Estimation

The final stage of our approach is the training of two separate single-value regression networks, one for the estimation of the patient’s height and the other for the weight.

In this stage, we attach untrained decoder heads to the pre-trained encoder bases. During the training we allow *all* parameters of the model, including the pre-trained encoder, to be fine-tuned. The motivation for a complete fine-tuning comes from two primary reasons: 1) pre-training of the encoder part allows the network to start with already good low-level features, so it is less-likely to turn these good low-level features to degenerate features during the fine-tuning, and 2) by still allowing changes in the low-level features we have the potential to squeeze further performance from the networks.

4 Experiments

4.1 Dataset

For training and validation of our method we have collected a very large dataset of 1899 volunteers and/or patients captured in 7620 clinical workflows, corresponding to nearly 170k depth images. Within this large dataset, we did not have the patient table information for 909 patients and the corresponding 909 workflows, so this subset was used only for the training of the height estimation network. Some example depth snapshots of this extensive dataset is provided in Fig. 2.

This dataset is collected from multiple sites in multiple countries, over a span of several years. The target clinical workflows such as a variety of coils, a variety of positioning equipment, unrestricted covers (light and heavy blankets), and occlusions by technicians, are covered in this dataset. Due to volunteer and patient consents, this dataset cannot be made publicly available.

We consider the following inclusion criteria for the training and validation: patient weight between 45 kg to 120 kg, patient height between 140 cm to 200 cm, and patient body mass index (BMI) between 18.5 to 34.9. The distribution of the samples in our dataset, together with the above inclusion criteria, is illustrated in Fig. 4.

4.2 Training

Training of the feature extraction network is performed using a subset of nearly 2000 workflows. For these workflows, we also acquired the corresponding full-body 3D medical volumes (MRI acquisitions). A set of 10 anatomical landmarks corresponding to major joints (i.e. knees, elbows, shoulders, ankles, and wrists) are manually annotated by a group of experts in these 3D medical volumes. These 3D annotations are transferred to depth image coordinates for training using the known calibration information.

We use a modified version of ResNet [13] as our base feature extraction network. This is a smaller version compared to the originally proposed ResNet18, where the number of features in each block is kept constant at 32. We use only bottleneck blocks instead of the basic blocks used for the original ResNet18.

Training of the feature extraction network is done using the ADAM optimizer [14] with default parameters. Landmark locations are represented as 2D



Fig. 4. Dataset distribution. Inclusion criteria is illustrated by the yellow region. (Color figure online)

heatmaps and mean squared error (MSE) loss is used. The feature extraction network is trained for 30 epochs with a patience of 10 epochs.

Once the feature extraction network is trained, we retain the encoder part and attach task-specific heads to form two separate networks, one for patient height estimation and the other one for patient weight estimation. Similar to the feature extraction network, we also train these networks using the ADAM optimizer with default parameters. As the loss function we use the symmetric mean absolute percentage error (SMAPE):

$$\text{SMAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|P_t - A_t|}{(|P_t| + |A_t|)/2} \quad (1)$$

where A_t is the actual value and P_t is the predicted value. We train for 250 epochs with a relatively large patience of 50 epochs.

For height estimation we omitted the second step of depth normalization through table subtraction since the patient back-surface does not affect the height estimation significantly.

4.3 Results

We evaluate our model using 23-fold cross-validation. Since we have multiple workflows and depth images corresponding to the same volunteer or patient (e.g. same volunteer captured both in a “knee-scan” acquisition and a “hip-scan”

Table 1. Patient height and weight estimation quantitative results.

PH5	PH15	95-Percentile Height Error
98.4%	99.9%	3.4%
<hr/>		
PW10	PW20	95-Percentile Weight Error
95.6%	99.8%	9.6%

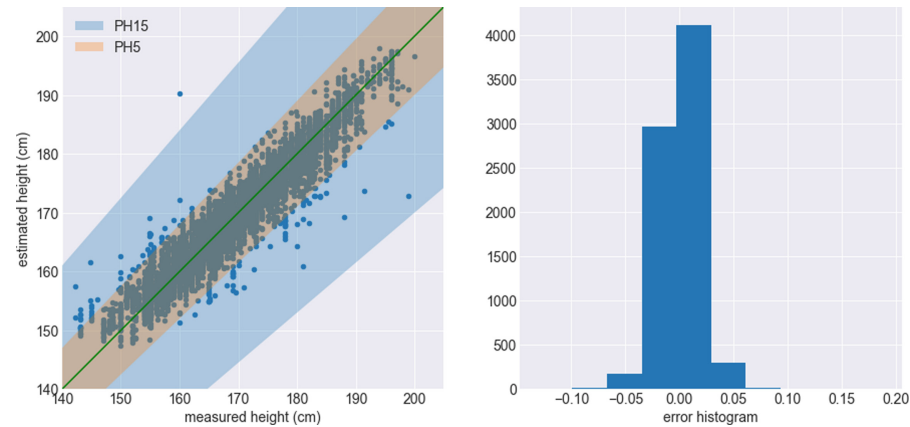


Fig. 5. Scatter plot and the error histogram for height estimation

acquisition), we ensure the splits are done based on the volunteer or patient identities. If a particular workflow contains more than one depth image, we apply a median filter on the frame-by-frame results to determine the final estimates for that workflow.

Table 1 provides our quantitative results. Our method achieved a PH5 of 98.4% and a PH15 of 99.9% for height estimation, and a PW10 of 95.6% and a PW20 of 99.8% for weight estimation, making the proposed method state-of-the-art in clinical setting. Figures 5 and 6 show the estimation scatter plots and the corresponding error histograms for height and weight estimation, respectively.

We also investigated the performance of our method for weight estimation for patient BMI groups outside our inclusion criteria. For patients with BMI < 18.5, our method achieved a PW10 of 89.2% and a PW20 of 98.9%. For patients with BMI > 34.9, our method achieved a PW10 of 96.2% and a PW20 of 99.8%. Even though the performance drops a little bit for underweight population, the main reason for keeping these populations outside the inclusion criteria is not the performance, but rather the limited support in the training dataset.

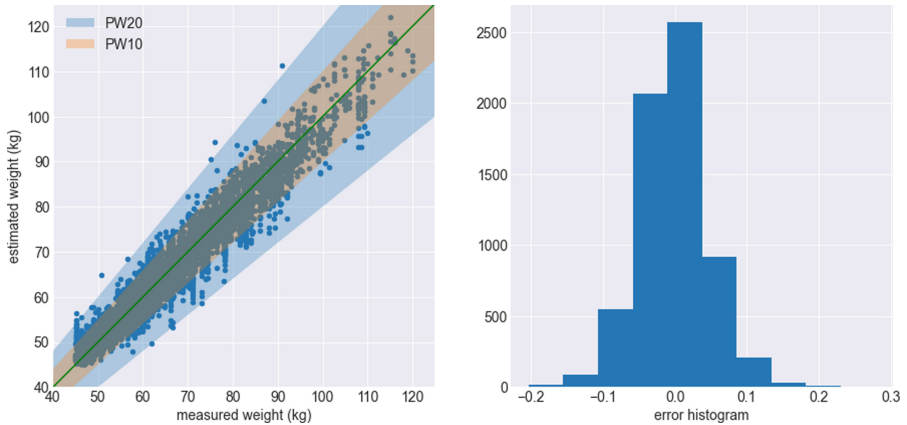


Fig. 6. Scatter plot and the error histogram for weight estimation

5 Conclusions

Accurate and rapid estimation of the patient's height and weight in clinical imaging workflows is essential for both the patient's safety and the possible patient-specific optimizations of the acquisition. In this paper, we present a deep-learning based method trained on a very large dataset of volunteer and patient depth images captured in unrestricted clinical workflows. Our method achieves a PH5 of 98.4% and a PH15 of 99.9% for height estimation, and a PW10 of 95.6% and a PW20 of 99.8% for weight estimation. These results out-perform all alternative methods to the best of our knowledge, including family estimates and clinical staff estimates.

Disclaimer. The concepts and information presented in this paper are based on research results that are not commercially available. Future availability cannot be guaranteed.

References

1. Lorenz, M.W., Graf, M., Henke, C., et al.: Anthropometric approximation of body weight in unresponsive stroke patients. *J. Neurol. Neurosurg. Psychiatry* **78**, 1331–1336 (2007)
2. Crandall, C.S., Gardner, S., Braude, D.A.: Estimation of total body weight in obese patients. *Air Med. J.* **28**, 139–145 (2007)
3. Wells, M., Goldstein, L.N., Bentley, A.: The accuracy of emergency weight estimation systems in children - a systematic review and meta-analysis. *Int J. Emerg. Med.* **10**, 1–43 (2017)
4. Abdel-Rahman, S.M., Ridge, A.L.: An improved pediatric weight estimation strategy. *Open Medical Devices J.* **4**, 87–97 (2012)
5. Wells, M.: A validation of the PAWPER XL-MAC tape for total body weight estimation in preschool children from low- and middle-income countries. *PLoS One.* **14**, e0210332 (2019)

6. Weber-Sanchez, A., Ortega Sofia, V., Weber-Alvarez, P.: Validation of the Broca index as the most practical method to calculate the ideal body weight. *J. Clin. Invest. Stud.* **1**, 1–4 (2018)
7. Buckley, R.G., Stehman, C.R., Dos Santos, F.L., et al.: Bedside method to estimate actual body weight in the emergency department. *J. Emerg. Med.* **42**, 100–104 (2012)
8. Kokong, D.D., Pam, I.C., Zoakah, A.I., Danbauchi, S.S., Mador, E.S., Mandong, B.M.: Estimation of weight in adults from height: a novel option for a quick bedside technique. *Int. J. Emerg. Med.* **11**(1), 1–9 (2018). <https://doi.org/10.1186/s12245-018-0212-9>
9. Akinola, O., Wells, M., Parris, P., Goldstein, L.N.: Are adults just big kids? Can the newer paediatric weight estimation systems be used in adults? *S. Afr. Med. J.* **111**, 166–170 (2021)
10. Cattermole, G.N., Wells, M.: Comparison of adult weight estimation methods for use during emergency medical care. *J. Am. Coll. Emerg. Physicians Open* **2**, e12515 (2021)
11. Pfitzner, C., May, S., Nüchter, A.: Body weight estimation for dose-finding and health monitoring of lying, standing and walking patients based on RGB-D data. *Sensors* **18**, 1311 (2018)
12. Dane, B., Singh, V., Nazarian, M., O'Donnell, T., Liu, S., Kapoor, A., Megibow, A.: Prediction of patient height and weight with a 3-dimensional camera. *J. Comput. Assist. Tomogr.* **45**, 427–430 (2021)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: *International Conference on Learning Representations (ICLR)* (2015)
15. Menon, S., Kelly, A.M.: How accurate is weight estimation in the emergency department? *Emerg. Med. Australas.* **17**(2), 113–116 (2005)
16. Fernandes, C.M., Clark, S., Price, A., Innes, G.: How accurately do we estimate patients' weight in emergency departments? *Can. Fam. Physician Medecin Famille Can.* **45**, 2373 (1999)