Estimating Body Fat from Depth Images: Hand-Crafted Features vs Convolutional Neural Networks

Marco CARLETTI[1], Marco CRISTANI[1], Valentina CAVEDON[2],
Chiara MILANESE[2], Carlo ZANCANARO[2], Andrea GIACHETTI[1]

[1] Department of Computer Science, University of Verona, Italy;

[2] Laboratory of Anthropometry & Body Composition, Dpt. of Neurological and Movement Sciences, University of Verona, Italy

Abstract

In this paper, we compare different approaches to estimate body fat percentages from simple depth images that can be captured by low-cost sensors. We implemented two frameworks, one based on hand-crafted features, using simple image processing methods to estimate directly from images a set of body measurements (e.g. areas, lengths girths), and one based on Convolutional Neural Networks, applying a direct regression from the grayscale maps representing the body depth, based on a pre-trained networks.

With these frameworks, we evaluated the fat percentage predictions obtained with the different methods on depth images of 350 subjects with known body composition estimated with a DXA scanner. Depth images were generated by extracting the z-buffer from the renderings of the 3D body scan models acquired on the group of subjects.

In our validation experiments, we evaluated the effect of different simulated acquisition setups, parameters settings, different image preprocessing and data-augmentation procedures and the addition of priors on height and weight on the prediction accuracy.

Furthermore, since the dataset used is composed of professional sportsmen and a control group, we evaluated also the ability of both frameworks of predicting the sport practiced by the subjects with a cross-validation experiment.

In specific, we propose a customized ResNet50 regressor to evaluate the whole body fat percentage of the subjects directly from the depth acquisitions. Using the same input data, we also set up a neural classifier to predict the sport category of the athletes.

Despite the limited numbers of subjects and the restricted variability of body types (all males, Caucasian, with a small number of obese), the results obtained are promising and can be considered a first step towards the development of quick and cheap body fat estimation tools that can be extremely useful for sport, health and fitness applications.

Keywords: 3d body scanning, body fat estimation, applications
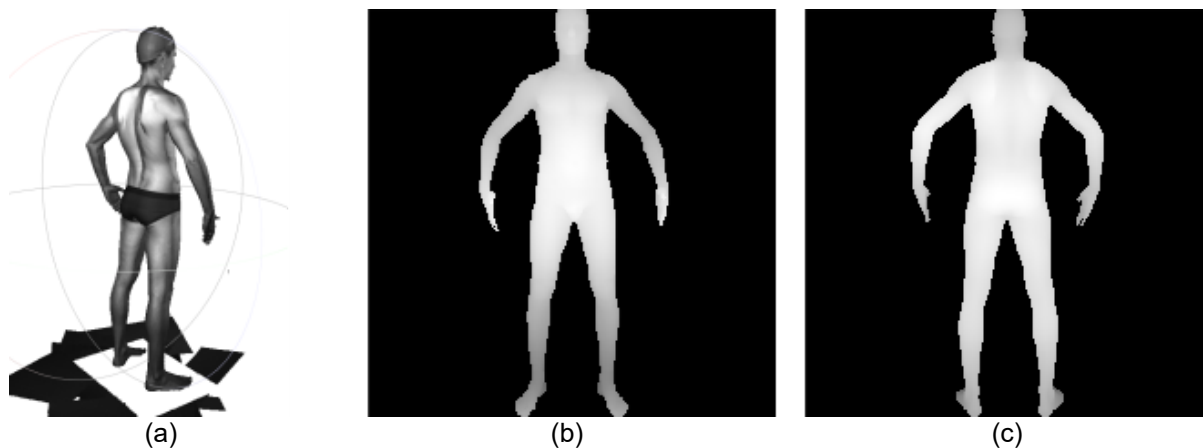
Figure 1: Original body scan (a) and rendered depth images used for fat regression obtained with different camera positions, (b) front view and (c) back view.

1. Introduction

Reasonable fat estimations are based on regression over anthropometric measurements [1, 2], which are commonly used as a method for body composition estimation. Sets of girths and lengths have been used for body fat estimation in several practical contexts, and well known formulas have been used for this purpose [3].

Manual anthropometry is, however, time consuming and prone to large inter-operator errors. A possible alternative is to use digital anthropometry, deriving the set of measurements from whole body 3D scans, [4, 5]. 3D scanning solutions designed for fitness and healthcare applications provide now specialized software packages providing body composition values assessed using regression formulas based on automatic measurements [6]. Accurate whole body scanning, however, is a time consuming task as well and requires expensive hardware to be performed.

The goal of this work is to estimate the fat percentage value from depth images, that one can acquire through cheap and fast sensors. Depth cameras, like Asus Xtion pro [8] or Intel Realsense [7], are quite cheap devices (approximately 100 Euros) providing a coarse 3D reconstruction of real world objects with relatively low resolution (>0.5cm). Considering, however, the large error (~1cm) in anthropometric landmark locations, it may be considered that the accuracy of measurement performed with a coarse depth sensor may be comparable with a manual measurement. Furthermore, using advanced Pattern Recognition techniques, like Convolutional Neural Networks, it may be possible to skip the measurement process and to link directly depth images with fat estimation, using a direct training procedure.

Estimating anthropometric parameters from depth maps is at its infancy, with few related works. Nguien et al. [9] estimate body weight from RGB-D images using regression on hand crafted features; Pfitzner et al. [10] add thermal camera data. These methods, however, are tested on a small number of subjects, and no publicly available data.

In our work, we investigated the possibility of performing image based body fat percentage estimation both using handcrafted features based on simplified body measurements and CNN based direct fat regression. In particular, our framework is based on a customized ResNet-50 [11] regressor, trained and tested to estimate fat percentage values directly from the front/back scans, achieving promising accuracy (standard errors of estimate SEE less than 2.1 on the depth renderings and 2.5 on a small set of real depth scans). Our choice is motivated by the fact that, using a convolutional neural network, there is no need to find accurate body landmarks: the regression is performed directly from the input images.

The paper is organized as follow: Section 2 presents the novel dataset used in the experiments, Section 3 describes the techniques chosen for the comparison. Section 4 presents the experiments results then discussed in Section 5.

2. Annotated depth images dataset

To analyze performances of body fat regression from depth images, we use a dataset of depth images rendered from 3D scans collected at the Department of Neuroscience, Biomedicine and Movement (DNBM) of Verona. As these body scans are collected with associated DXA acquisition data and anthropometric measurements, we can associate to each model and to rendered images both fat percentages and weight and height of the subjects. Examples of the rendered images are shown in Figure 1. We rendered four versions of the simulated depth images, two representing the whole body from front and back view, and two representing only the roughly segmented torso area.

We collected a total o 350 subjects, all males, 250 are professional sportsmen and 100 are a control group. The average age of the group is 26.5 (7.8 std.), and the whole body fat percentage spans from 6.64 to 36.78. The distribution of the whole body fat percentage, estimated with the DXA scanner in the group is portrayed in Figure 2. The distribution has a mode around 10%, with a long right tail. Depth images have been created simulating a front and a back acquisition from a distance of about 1.5 m with an Asus Xtion Pro sensor. All the images were obtained by performing depth renderings using Blender [12], where the engine options were set in order to simulate the Asus Xtion Pro depth sensor.

Images have been rescaled in order to fit the height of a fixed size image (224x224 pixels, as needed by the ResNet50 architecture), discarding the body size information. However, in our work we try to estimate whole body fat percentage, not an absolute value, so we can avoid the estimation of exact body and evaluate body shape features mapping the images to a reference scale. The information about actual body size can be then added putting the subjects' height as further input for the regression algorithm. Subjects' weight can be added as another feature for the regression as well.
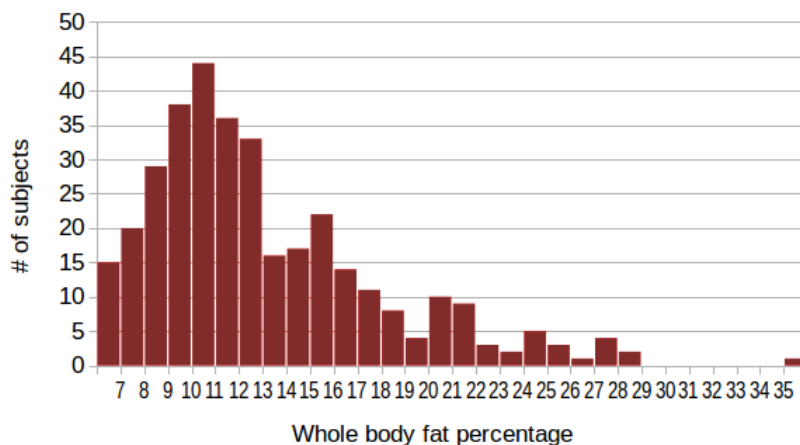

Figure 2: Distribution in terms of whole body fat percentage of the dataset.

3. Body fat estimation from depth images

3.1. Hand crafted features

A simple solution to try to estimate the fat value from the depth images is to try to perform a set of measurements on them, similarly to what is done in the 3D approaches based on geometry processing, and shown, for example, in [5]. It is, however, quite difficult to locate accurate landmarks on low resolution depth images, so, the selection of relevant features must be considered with care.

On full body rescaled images, we considered: the area of the silouhette, the apparent diameter of the neck, the maximal and minimal width of the trunk, the depth variation in the abdominal area, the depth variation in thoracic area. We tried to estimate and add to the regression framework other features, estimated on the trunk region or on back images, but these features did not improve the body fat accuracy obtained.

Figure 3 shows example results of the automatic mesh segmentation performed on three models of our dataset.
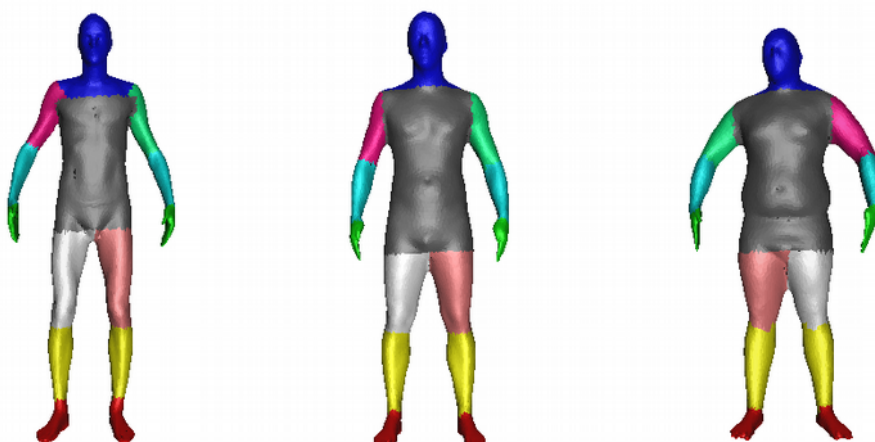

Figure 3: Rendered 3D scans from the athletes dataset, processed with the method described in [5] to estimate segment body parts and evaluate anthropometric measurements.

Measurements are put in a vector and regression algorithms are trained on sets of feature vectors with associated ground truth fat percentages.

As the simple depth acquisition can be easily integrated with height and weight measurement, and we have the data in our dataset, we also investigated the effect of adding these values to the feature vector.

3.2. Convolutional Neural Networks

Convolutional Neural Networks are quite popular Pattern Recognition tools that demonstrated an impressive ability of predicting categories and numerical values of imaged scene parameters directly from the pixel array input.

The idea of this work is to build a deep network linking image layers with convolutional filters adding nonlinear units and pooling steps and with parameters to be learned by minimizing a residual functions. The goal of our CNN-based fat estimator is to provide a method for automatically measure whole body fat percentage from a inexpensive and quick range scan. As the previously described dataset was composed of simulated depth scans, even if performed on real body geometries, and even if the rendering was made with a resolution matching the specifications of the current generation of cheap depth sensors, it would be clearly better to work on directly acquired depth images.

The ImageNet ResNet architecture [11] is well-known for performing extremely well in classification tasks. We convert a pretrained ResNet-50 model into a regressor by removing the last classification layer, substituting it with a fully connected one, mapping the convolutional features directly to the raw fat index. Two different regressors are trained on the Front data (providing $\overline{y}_{Front}$ estimates) and the Back version. The, the information of the regressors are fused together: we estimate the train error as the average of the standard error SEE $\varepsilon = \|\overline{y} - y\|$ on all the samples of each view, giving $\overline{\varepsilon}_{Front}$ and $\overline{\varepsilon}_{Back}$. These average errors are normalized and turned into weights:

$$w_{Front} = 1 - \frac{\overline{\varepsilon}_{Front}}{\overline{\varepsilon}_{Front} + \overline{\varepsilon}_{Back}}$$

and similarly for w_{Back} so that the combined regression on a given sample is computed as $w_{Front} * \hat{y}_{Front} + w_{Back} * \hat{y}_{Back}$. The results of this method and the classic measurement-based approach presented in the Subsection 3.1 are then compared.

4. Experimental results

In our validation tests, we performed 10-fold crossvalidation experiments in order to estimate fat percentage predictions for the whole dataset. We sorted the subjects according the ground truth fat and divided the set in 10 subsets with similar fat distributions. We then trained 10 regressors, each one with data of 9 subsets, estimating the values of the other subset in order to get unbiased estimate for the whole group of subjects.
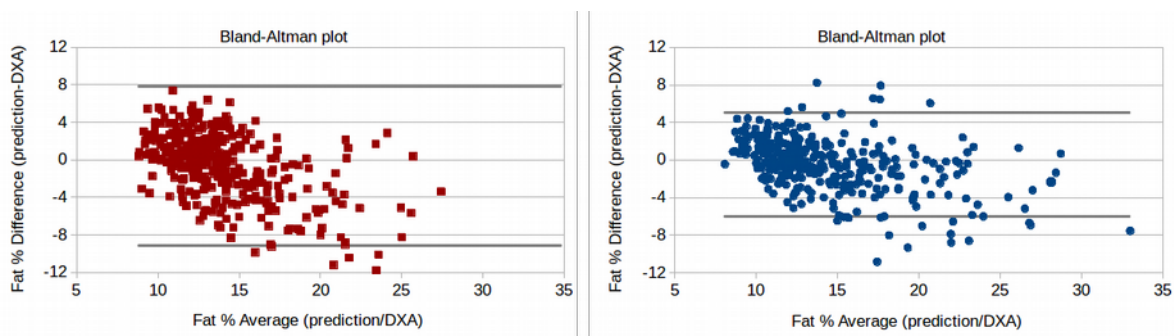

Figure 4: Bland Altman plots showing the relevant improvement in the 10-fold crossvalidation body fat estimation obtained by using CNN based regression (right) over simple hand-crafted features estimated on rough depth maps (left).

In each training phase, in order to address the fat index unbalance (already discussed in Sec. 2), we perform non-uniform data augmentation, taking the samples between 10 and 18 fat percentage values, augmenting them with noise injection and horizontal flipping with a rate of 2 with respect to the inlier samples. Then, we obtain fat percentage estimates for all the subjects in the database and all the methods, exploiting maximally the limited number of subject data and trying to compensate the fat distribution unbalance.

In Figure 4, the Bland-Altman plots portraying for the Full-Body and Torso versions, the distribution of the differences between the estimates and the reference DXA values of the fat percentage are shown. From these graphs, it worth noting that the estimates are not biased and the average error is constant across the fat range with only few outliers lying outside the lines of agreement.

Table 1 shows the standard error of estimates and pearson correlation obtained with 3D models and the measurement set presented in [5], depth images with handcrafted features and CNN-based regression, performed on front and back rendering and also on renderings of roughly segmented torso area. As visible, error is lower on Torso data: this can be explained by the higher xy resolution of the data (e.g. body shape encoding) and the limited effect of pose variations. Combined regression gives a better performance, since more information is jointly considered. In general, the error is reasonable and actually lower than the one obtained with the measure-based framework in [5]. Although further analysis have to be performed, our depth image based fat regression seems reasonably robust against the variation introduced by the arms and legs position.

The addition of the weight and height values does not seem to significantly improve the fat percentage estimate in the 3D case, while it is effective in the 2D hand-crafted and CNN-based methods.

Table 1
Standard Error of Estimates and Pearson coefficients for 10-fold crossvalidation regression

Method	SEE	Pearson
2D Hc	3.24	0.52
2D Hc + WH	3.04	0.55
3D Meas.	2.70	0.71
3D Meas + WH	2.68	0.71
CNN Front (full)	2.34	0.79
CNN Back (full)	2.35	0.78
CNN Front (torso)	2.18	0.82
CNN Back (torso)	2.24	0.80
CNN Back+Front (full)	2.20	0.80
CNN Back+Front (torso)	2.05	0.83
CNN Back+Front + WH	1.87	0.83

We also tried to automatically predict, with a similar 10-fold crossvalidation scheme, the subjects' practiced sports, both with hand-crafted feature sets estimated on 3D models, 2d depth images and CNNs. The average classification accuracy obtained is 41.8 with 2D features, 48.9 with 3D features and support vector machine classifier and 55.5 with the ResNet50 based classifier.

5. Discussion

Our work shows that it is possible to design cheap estimators of body fat percentage based on simple depth sensor data, obtaining reasonably good accuracy. The use of CNN based regression using the whole depth image information, robustly against spatial transformation seems fundamental to ensure a good results. We were not able to obtain similar accuracy using hand-crafted features, even working directly on the 3D body scans. The errors obtained on our dataset are reasonably small (smaller than those obtained with a simple measurement-based techniques) and the early results on low cost depth sensor data are encouraging.

However, a mandatory step to will be the creation of larger datasets including non-fit subjects with wider fat variation, trying to balance the dataset and reaching an order of magnitude more in terms of number of samples. To develop a generic application for body fat estimation from depth sensor data, it is necessary to train the system with images of both men and women and subjects belonging to different ethnic groups as the fat distribution patterns varies within the different subset.

We believe that the proposed SportsDepths dataset is quite interesting for research purposes and we decided to make it available for the scientific community. Depth images are available at the web address www.andreagiachetti.it/SportsDepths, with associated metadata, including fat percentages and the label related to the sport practiced by each subject. The reasons why we consider the dataset interesting are the following: first, using this public benchmark, CNN-based regression methods can be tested and improved, and also fat estimation application can benefit of these improvements; second, a depth based fat estimation application represents for sure an attractive commercial perspective: cheap stereo sensors are invading the market, while other inexpensive solutions of fat estimation based on bioelectric impedance, skinfold or anthropometric measurements are not reliable or require time consuming and/or expensive procedures.

References

[1] W. E. Siri. The gross composition of the body. Adv Biol Med Phys, 4(239-279):513, 1956

[2] V. H. Heyward, D. R. Wagner, et al. Applied body composition assessment. Number Ed. 2. Human Kinetics, 2004

[3] Standards of medical fitness. Washington, DC:Department of the Army, 4, 2011

[4] T. N. Garlie, J. P. Obusek, B. D. Corner, and E. J. Zambraski. Comparison of body fat estimates using 3d digital laser scans, direct manual anthropometry, and dxa in men. American Journal of Human Biology

[5] A. Giachetti, C. Lovato, F. Piscitelli, C. Milanese, and C. Zancanaro. Robust automatic measurement of 3d scanned models for the human body fat estimation. IEEE journal of biomedical and health informatics, 19(2):660–667, 2015

[6] Styku web site: http://www.styku.com/ / last visited 18/8/2018

[7] Intel Realsense, web site: https://software.intel.com/en-us/realsense last visited 18/8/2018

[8] Asus Xtion 2, web site: https://www.asus.com/3D-Sensor/Xtion-2/ last visited 18/8/2018

[9] T. V. Nguyen, J. Feng, and S. Yan. Seeing human weight from a single rgb-d image. JCST, 29(5):777–784, 2014

[10] C. Pfitzner, S. May, and A. Nüchter. Neural network-based visual body weight estimation for drug dosage finding. In Medical Imaging: Image Processing, page 97841Z, 2016

[11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In Proceedings of CVPR, pages 770–778, 2016

[12] Blender Online Community. Blender - a 3D modelling and rendering package. Blender Foundation, Blender Institute, Amsterdam, 2017