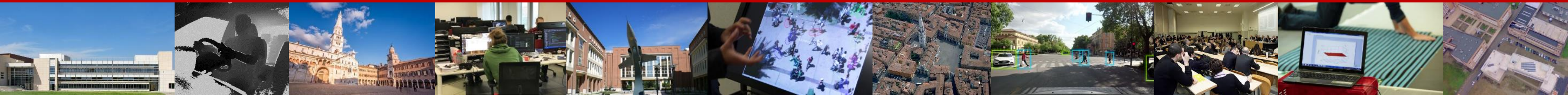
IJCB 2020

Baracca: a Multimodal Dataset for Anthropometric Measurements in Automotive



Stefano Pini, Andrea D'Eusanio, Guido Borghi, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

softech-ict Centro Interdipartimentale di Ricerca Softech: ICT per le Imprese

The Baracca Dataset

A Multimodal Dataset for Anthropometric Measurements in Automotive

- >9k frames in 4 modalities
- 2 synchronized cameras (RGB-D, Thermal*)
- 4 streams: RGB, IR, Depth, Thermal
- 8 viewpoints (3 in-car, 5 outside)
- 30 subjects
- Anthropometric measurements:
  height, shoulder width, forearm and arm length,
  torso width, leg length, eye height from the ground.
- Soft-biometric traits: age, sex, weight
- Human keypoints estimated with HRNet[1]
- Available at https://aimagelab.ing.unimore.it/go/baracca

	Height	Eye Height	Forearm	Arm	Shoulders	Torso	Leg	Age	Weight	BMI
Mean	175.2	164.6	25.73	26.67	42.27	38.63	103.8	26.57	72.03	23.35
Std. dev.	7.100	7.059	1.879	2.134	3.255	2.702	5.536	3.981	12.71	3.222

**. Radiometric LWIR Thermal Camera*

1. Sun et al., Deep high-resolution representation learning for human pose estimation. CVPR 2019

The Baracca Dataset

Acquisition Setting

Goals:

- Automotive scenario
- Realism / Applicable to real-world systems/cars
- Enabling
  - Anthropometric measurements
  - Soft biometrics estimation
  - Face/person recognition

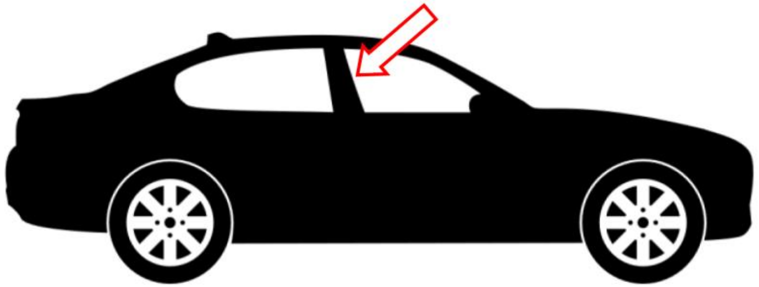
The Baracca Dataset

Acquisition Setting

Two scenarios:

- **Outside view**
  The cameras can be placed:
  - On the B pillar (external side)
  - With variable tilt
  - 5 different sensor-subject distances

- **In-car view**
  The cameras can be placed in different positions:
  - A pillar
  - Rear mirror
  - Behind the steering wheel



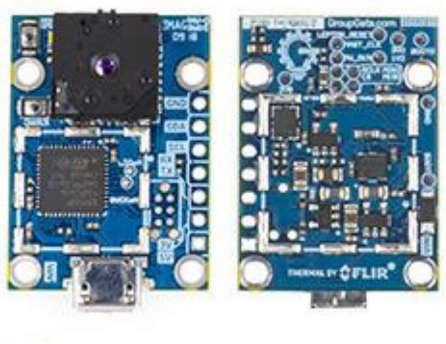
The Baracca Dataset

Acquisition Devices

Two synchronized devices (four streams):

- **Pico Zense DCAM710**[1] provides:
  RGB: 1920x1080px
  IR: 640x480px (16bit)
  Depth: 640x480px (16bit, millimeters)

- **PureThermal 2 with Flir Letpon 3.5**[2] provides:
  Thermal: 160x120px (radiometric, 16bit)



1. *https://www.picozense.com/*
2. *https://store.groupgets.com/products/flir-lepton-3-5* + *https://store.groupgets.com/collections/frontpage/products/purethermal-2*

The Baracca Dataset

Acquisition Procedure

- Recording of visual data:
  - Acquisition of 10 synchronized images for each viewpoint
  - Out-of-car acquisitions
    - Different viewpoints at known distances: 0.6m (top-view), 0.6m, 1m, 1.5m, 2m
  - In-Car acquisitions
    - Different viewpoints: left A pillar, rear mirror, behind the steering wheel
- Collection of soft-biometric traits:
  - age, sex and weight
- Collection of anthropometric measurements:
  - height, shoulder width, forearm length, arm length,
    torso width, leg length, eye height from ground

Anthropometric measurement estimation

We explored three different baseline approaches:

- A purely-geometric approach
- A set of machine learning-based approaches
- A deep learning-based approach

We evaluated them on:

- Several data types
- Different viewpoints

Anthropometric measurement estimation

Geometric Approach

This method geometrically estimates the distance between the head of the person and the ground and between the eyes and the ground.

Input: depth map, camera calibration parameters

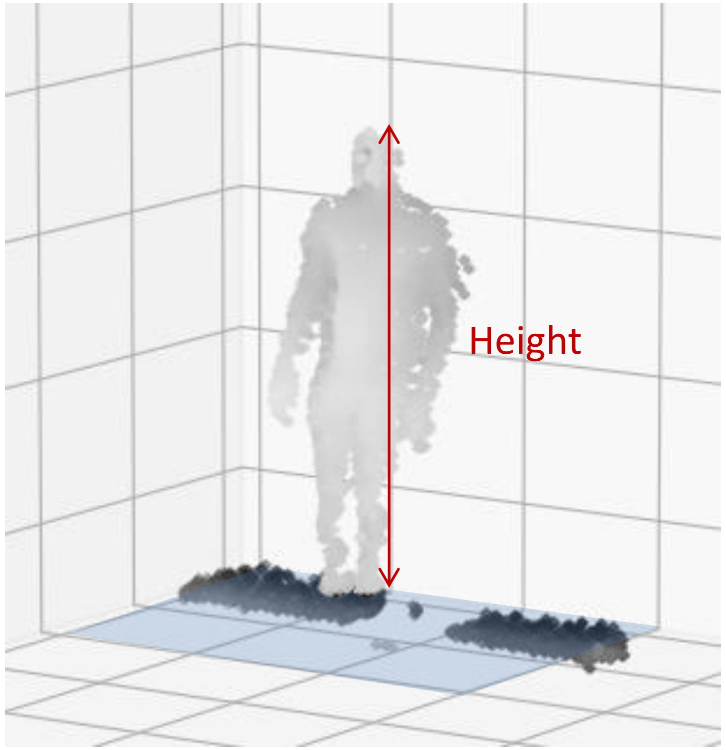
Output: height, eyes height from ground

Method:

- Depth map is converted to a point cloud
- RANSAC algorithm is used to estimate the plane corresponding to the ground
- A point-to-plane distance is calculated to retrieve the height and the eye height of the subject

Key elements:

- It only works on the depth data of the outside viewpoints
- It does not require any training.



Anthropometric measurement estimation

Machine Learning Approaches

This set of machine learning-based methods convert a whole set of joint lengths to the desired anthropometric measurements.

Input: a set of joint lengths, in pixels (2D) or mm (3D)

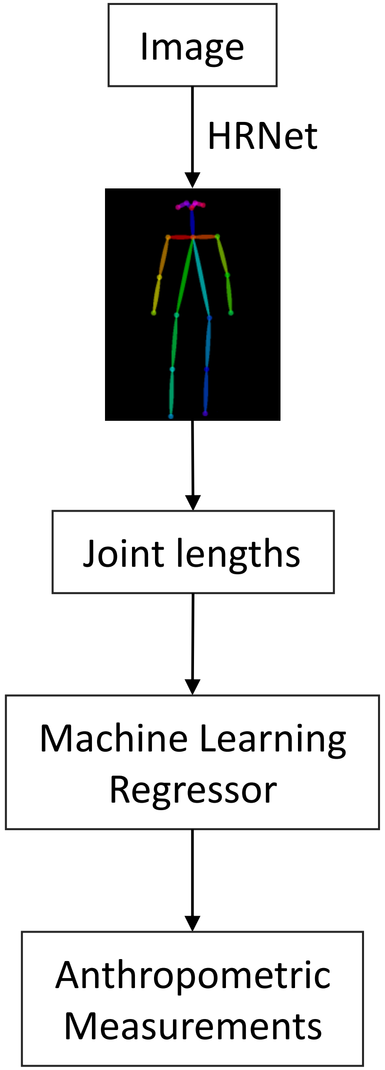
Output: anthropometric estimations

Methods:

- Linear Regression – least-squares linear model
- Random Forests – ensemble of regression decision trees
- AdaBoost – combination of weak regressors

Key elements:

- Does not directly use images as input
- Rely on HRNet[1] body joint predictions



1. Sun et al., Deep high-resolution representation learning for human pose estimation. CVPR 2019

Anthropometric measurement estimation

Deep Learning Approach

This deep learning-base model directly estimates the anthropometric measurements from the visual data.

Input: a single image of a specific type (RGB, IR, Depth, or Thermal)

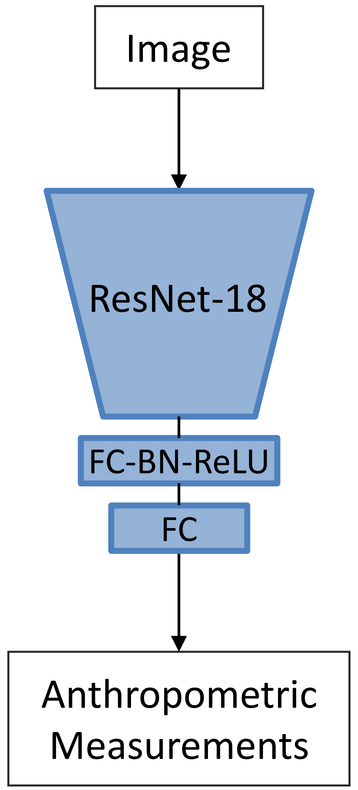
Output: anthropometric estimations

Method:

- ResNet-18[1], pre-trained on ImageNet[2], is used as feature extractor
- Features are processed by two fully connected layers
- The last fc layer regresses the anthropometric measurements

Key elements:

- The input is just an image
- The Huber Loss[3] (δ=1) is used as objective function



1. *He et al., Deep residual learning for image recognition. CVPR 2016*
2. *Deng et al., Imagenet: A large-scale hierarchical image database. CVPR 2009*
3. *Huber., Robust estimation of a location parameter. Breakthroughs in statistics, Springer, 1992*

Result summary

- Results (reported in the paper) show that anthropometric measurements and soft-biometric traits can be successfully estimated using any data type included in the dataset.

- The machine learning approaches outperform the other methods, but they rely on the body joint predictions of HRNet.

- 3D data are shown to be the most suitable data for the anthropometric estimation. In view of this:
  - depth sensors are the most adequate sensors for anthropometric estimation.
  - 3D data can be also roughly estimated from 2D sensors only if
    - the distance between the subject and the camera is known
    - the subject joints can all be considered at the same distance.

Result summary

	Pros	Cons
Geometric Approach	• 3D-based, understandable • Works even with extremely-small point clouds	• Requires a depth sensor • Less precise than other methods • Slow
Machine Learning Approaches	• Accurate, best-performing solution • Work with both 2D and 3D data (at known distance)	• Subjects must be at a known distance when using 2D sensors • Rely on predictions of a Human Pose Estimation method (HRNet)
Deep Learning Approach	• Faster than other methods • Good tradeoff between speed and accuracy • Additional soft-biometric traits can be estimated	• Subjects must be at a known distance when using 2D sensors • Prone to overfit on the training set and the training subjects

IJCB 2020

Thank you for your attention

Baracca: a Multimodal Dataset for Anthropometric Measurements in Automotive

Dataset available at https://aimagelab.ing.unimore.it/go/baracca

Stefano Pini, Andrea D’Eusanio, Guido Borghi, Roberto Vezzani, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

Centro Interdipartimentale di Ricerca Softech: ICT per le Imprese