Accurate Human-Limb Segmentation in RGB-D images for Intelligent Mobility Assistance Robots

Siddhartha Chandra, Stavros Tsogkas & Iasonas Kokkinos
siddhartha.chandra@inria.fr, stavros.tsogkas@centralesupelec.fr, iasonas.kokkinos@inria.fr

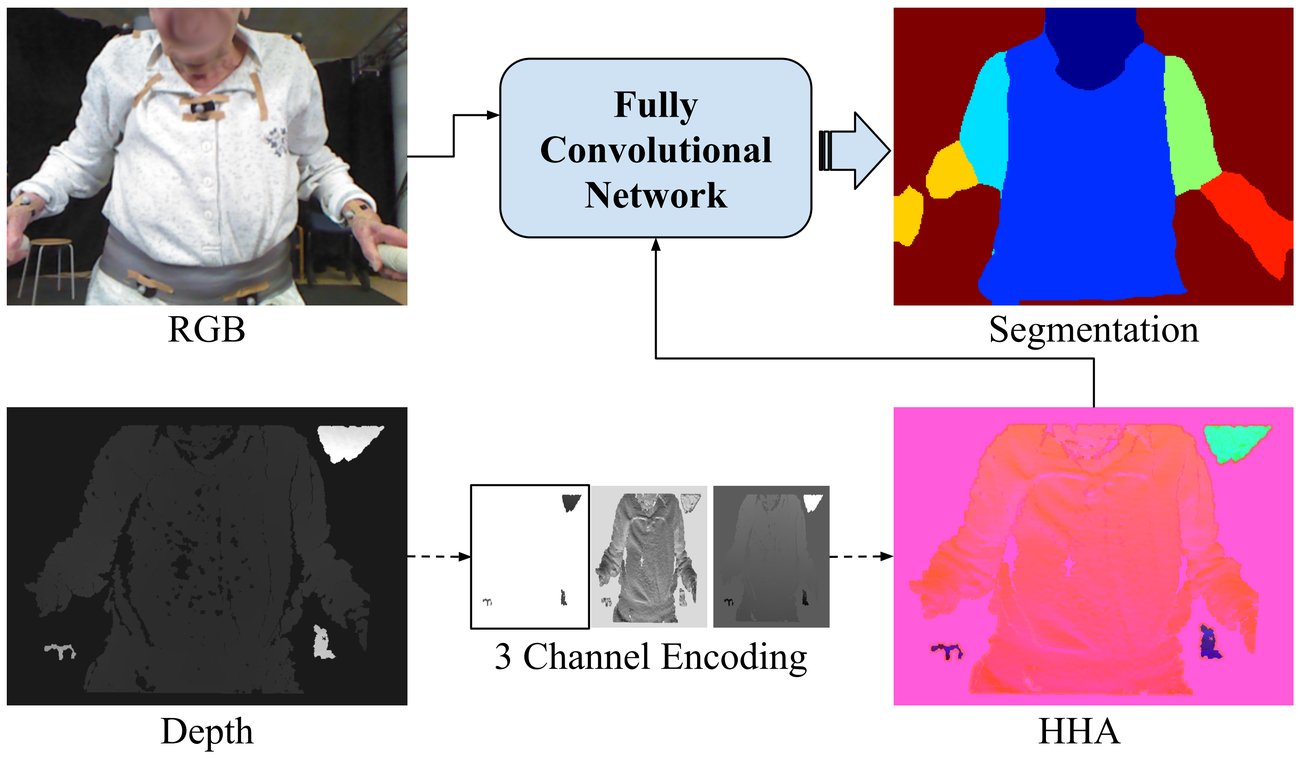
Motivation

- Mobility impairment widespread among the elderly people
  - 20% of people between the age of $70-85$ years
  - 50% of people above the age of 85 years
- EU project developing an intelligent robot for elderly user assistance

Contributions

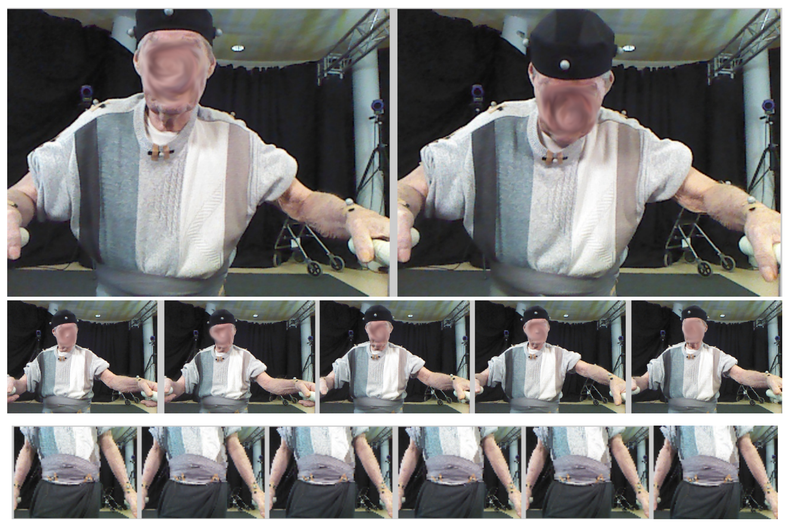
- Deep-learning for accurate semantic segmentation of human limbs using colour and depth images.
- Scheme for manual annotation of videos, that eliminates the need to annotate segmentation masks in every single frame
- Extend state of the art deep learning systems for semantic segmentation, object detection to exploit diverse RGB and depth data, in a single framework for training and testing
- Promising performance on our in-house Human-Limb dataset

Deep Learning Pipeline



- Extend and combine state of the art methods
  - Deeplab [2] network for semantic segmentation in RGB images
  - Alexnet-HHA [1] network for object detection using depth image
- Concatenate penultimate layers of RGB, HHA networks
- Single loss layer on top
- Dense CRF for post processing network output
- Use of diverse data

Clustering For Annotation



- Clustering based strategy to avoid annotating very similar frames
- Clustering metric is similarity of images in the HOG space
- Annotate one image per cluster
- Propagate annotation of representative image to others

Human-Limb Dataset

2 example image-sets from the dataset. Each block shows the RGB image, the depth image, the depth image encoded to 3 channel HHA [1], and the pixelwise ground truth class labels.

References

[1] S. Gupta, R. Girshick, P. Arbeláez, J. Malik Learning rich features from RGB-D images for object detection and segmentation In *ECCV*, 2014.

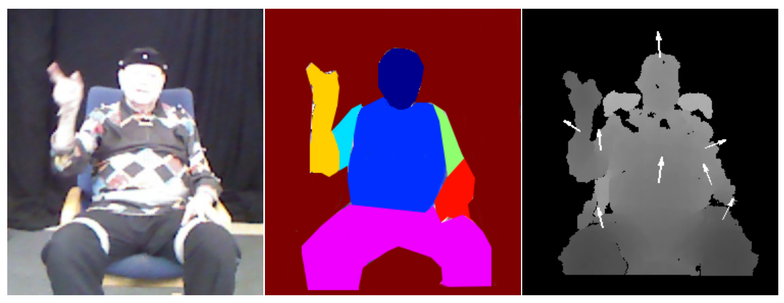
[2] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille Semantic image segmentation with deep convolutional nets and fully connected crfs arXiv, 2014

Quantitative Results

Method	Softmax	Dense CRF
RGB	86.62	87.26
HHA	83.09	84.94
RGB+HHA	86.99	87.45
(RGB + HHA) + INRIA	**88.12**	**88.84**

- Pixelwise segmentation accuracy
- Four types of training data
- Additional negatives (INRIA)
- Dense CRF postprocessing

Human Stability Application

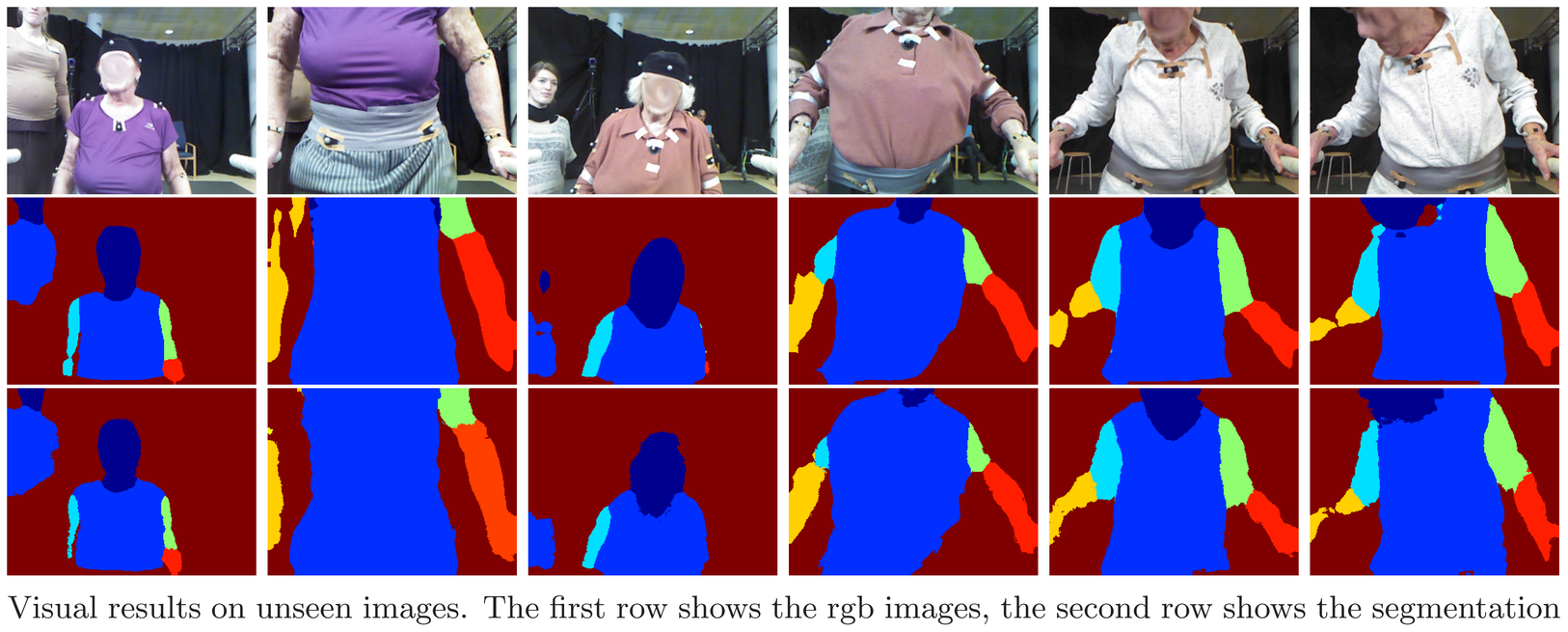


- Segmentations and depth for fitting surfaces to body parts
- Surface normals indicate pose stability
- An unstable pose alerts the robot
- Robot uses normals to estimate the magnitude, direction of optimal force

Summary

- Scheme for manual annotation of videos that avoids redundant work
- Deep learning solution to human limb segmentation in RGB-D images that exploits existing state of the art methods
- Efficient implementation built on top of *Caffe*
  - Testing speed: 8 fps
  - Allows learning seamlessly from diverse data: RGB, D, or RGB-D images
- Promising results on our in house human limb dataset
- Applications to detection, gesture and pose estimation, and stability estimation

Visual Results



Visual results on unseen images. The first row shows the rgb images, the second row shows the segmentation results of our network, and the third row shows the results after using dense CRF.