Aarila Dots Oy
09.09.2019

v 1.0.0

Study of the automatic garment measurement

REPORT

European Union
European Regional
Development Fund

Leverage from
the EU
2014–2020

SATAKUNTALIITTO
Regional Council of Satakunta

Prizztech

1. Description of this study

1.1 Introduction

The measurement of a garment is a crucial factor in the garment making industry when maintaining a consistent size grading. Due to the qualities of the fabric, even automated cutting and stitching have some variations in the product dimensions. Revising the measures is used as a form of quality control where garments with too deviating measures get rejected, or their grading is updated.

Reuse and recycling industries can also benefit from the garment measurement. When garments arrive in the recycling center, they can have missing or dramatically inconsistent grading labels that cannot be relied on. If the items are resold, measuring the garment can provide valuable size assurance for the customer.

The traditional way of manually measuring the garment is labor and time-intensive, and precision is affected by human factors. Automated computer vision-based measurement solutions could provide more precise measures and add standardization to the process.

1.2 Overview

The objective of this study is to report the possible ways of measuring the garment automatically and examine the limitations to them. And also, to determine to what extent the recent advances in computer vision and deep learning can be used to advance the measuring technology.

1.3 Disclaimer

The contents of this study, including but not limited to the information about the regulations, laws, available services, functionality and features of services, limitations, restrictions, instructions and other advice have been provided by Aarila Dots Oy for informational purposes only and are not intended as and do not constitute legal advice or opinion. We have made efforts to ensure the accuracy of the information provided, however, we do not promise or guarantee that the information is correct, complete or up to date. Aarila Dots Oy expressly disclaims all liability in respect to actions taken or not taken based on any or all the contents of this study.

European Union
European Regional Development Fund
Leverage from the EU 2014–2020

1.4 Authors of the study

This study has been conducted by Aarila Dots and the study was led by the chief technology officer of Aarila Dots; Joni Juvonen. The extensive technical expertise of the rest of the Aarila Dots team was also utilized during this study.

The team members of Aarila Dots Oy have worked in various different computer vision and deep learning projects as well as have developed a computer vision mobile application for taking body measurements called TailorGuide, which links to service for getting garment size recommendations in apparel online stores.

In addition to various body size and garment related photogrammetry projects, the team has expertise in the fields of cloud platform building, biomedical imaging, digital pathology, and imaging-based quality control in manufacturing processes. It is notable, that the team members have placed in the top 3%, top 4%, and top 10% in Kaggle's international machine learning and data analysis competitions. The team also has proficiency in multiple programming languages and different software platforms including but not limited to; Python, Java, C#, Kotlin, Swift, JavaScript, OpenCV, TensorFlow, Keras, PyTorch, and FastAi.

2. Automatic garment measurement

The idea of measuring garments automatically is not new and methods have appeared in Stitch Fix photo-based clothing measurement blog post [1] as well as in a number of academic publications [2] [3] [4]. What they all share in common, is that they capture a garment that is laid flat on a surface with a calibrated camera. Some methods detect the clothing type and then the measurement points and some just the latter. Finally, they calculate the physical distances between key points converting

[1] Coffey B. and Torres T.J. 2016. Photo-Based Clothing Measurements https://multithreaded.stitchfix.com/blog/2016/09/30/photo-based-clothing-measurement/ [Accessed Aug. 13th, 2019]

[2] Chunxiao Li, Ying Xu, Yi Xiao, Huimin Liu, Meiling Feng, and Dongliang Zhang. 2017. Automatic Measurement of Garment Sizes Using Image Recognition. In Proceedings of the International Conference on Graphics and Signal Processing (ICGSP '17). ACM, New York, NY, USA, 30-34. DOI: https://doi.org/10.1145/3121360.3121382

[3] Li Cao, Yi Jiang, and Mingfeng Jiang. 2010. Automatic measurement of garment dimensions using machine vision, 2010 International Conference on Computer Application and System Modeling(ICCASM 2010), Taiyuan, 2010, pp. V9-30-V9-33. DOI: https://doi.org/10.1109/ICCASM.2010.5623093

[4] Kunchang Chen , 2005. Image Analysis Technology in the Automatic Measurement of Garment Dimensions. Asian Journal of Information Technology, 4: 832-834. DOI: http://medwelljournals.com/abstract/?doi=ajit.2005.832.834

from pixel values to millimeters or inches using the known calibration coefficients from the camera.

The process can be roughly divided into four steps:

1. Camera calibration
2. Clothing type classification
3. Measurement point detection
4. Physical distance calculation

2.1 Camera calibration

In this context, the camera calibration means mapping the pixel coordinates to a physical measurement plane of known dimensions. It includes correcting the radial and tangential distortion. The radial distortion is caused by the symmetrical shape of the lens and it increases as the distance from the focal center grows. Tangential distortion may be present if the lens is misaligned with the camera's sensor. The radial distortion can be corrected with equation 1 and the tangential with equation 2. This uses distortion coefficients $k_1,\ k_2,\ p_1,\ p_2,\ k_3$ and these intrinsic parameters are typically calculated from capturing a known chessboard or circular pattern and examining the curvature of straight lines.[5]

Equation 1

$$x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$
$$y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

Equation 2

$$x_{corrected} = x + [2p_1 xy + p_2(r^2 + 2x^2)]$$
$$y_{corrected} = y + [p_1(r^2 + 2y^2) + 2p_2 xy]$$

Other intrinsic parameters include the focal length and optical center of the sensor in pixel coordinates. Intrinsic parameters (distortion coefficients, focal length, and center) are specific to the camera and once these are known, they don't need to be recalculated unless the lens is changed.

The extrinsic parameters are needed for translating the pixel coordinates to the measurement plane's 3D coordinates or vice versa (Figure 1). These include the camera's rotation and translation with respect to the platform. Some of the garment measurement methods perform this part of the calibration only once although the

[5] OpenCV camera calibration and 3D reconstruction https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_calib3d/py_calibration/py_calibration.html [Accessed Aug. 13th, 2019]

extrinsic parameters can change due to shifting or tilting of the camera. A more robust way is to calibrate the extrinsic parameters in each capture by locating a known calibration pattern such as a square bounding rectangle from the measurement plane [1]. This will add a varying projection error that can change between the captures but the amount of error can be reliably measured. A good metric for this is the re-projection error, where the known 3D calibration pattern corner points are projected to the image plane with the calculated transformation parameters, and the L2-norm is calculated in respect to the edge detected pattern corners.

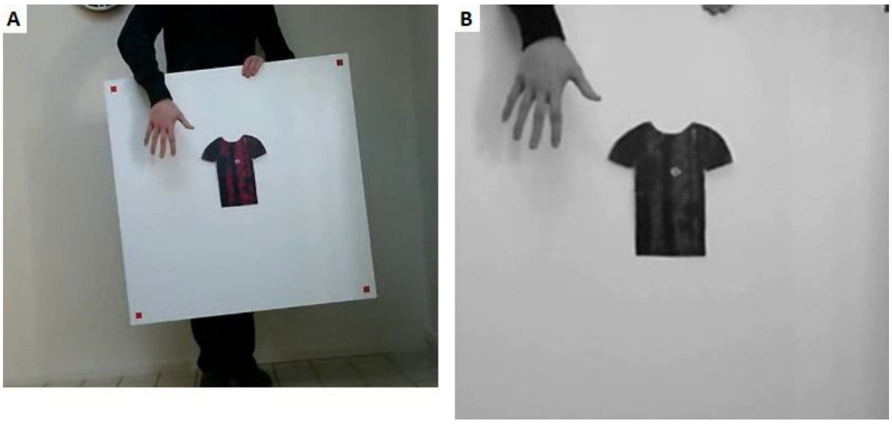

Figure 1. Platform corner markers are located from image A and projected to XY-plane (B) with perspective transformation.

2.2 Clothing type classification

Automatic clothing type classification is an optional step in garment measurement that is only relevant if multiple types such as t-shirt, shirt, pants, skirt, etc. are measured and no prior information of the type is provided manually. Li et al. (2017) [2] detected the type out of six (T-shirt, shirt, pants, vest, skirt, and one-piece dress) by performing template matching to the detected garment contours. Additionally, they divided each category into two subcategories to account for variations such as long and short-sleeved t-shirts.

Advances in deep learning techniques in computer vision and large labeled image datasets have advanced object classification and localization methods [6] [7] [8].

[6] ImageNet object localization challenge https://www.kaggle.com/c/imagenet-object-localization-challenge [Accessed Aug. 21st, 2019]

[7] COCO - Common objects in context http://cocodataset.org/ [Accessed Aug. 21st, 2019]

Convolutional neural networks (CNN) have emerged as the go-to method for complex tasks such as clothing type recognition from images [9]. These have the capacity to learn from large datasets and achieve better results than traditional computer vision methods. CNN's have even exceeded human-level performance in many image-related tasks and compared to traditional template matching they have additional advantages besides the performance [10]. CNN can learn the features of an object from examples and generalizes well given diverse training examples. A template matching algorithm detects only close matches to designed general-shape template patterns and does not perform well on objects that vary from these.

The performance of just the clothing type classification was not specified in the Li et al. (2017) [2] publication but this depends largely on the image train and test sets. Normalized filming conditions such as fixed camera location and good lighting remove some of the environmental variations that could complicate the detection. For getting a sense of the expected performance in clothing type classification, we trained a Single shot multibox detector (SSD)[11] with DenseNet121[12] base CNN architecture for garment object detection and classification. We used parts of the DeepFashion image dataset for training, validation, and testing, and selected a subset of the 18 most frequent clothing which were blazer, blouse, cardigan, dress, hoodie, jacket, jeans, joggers, jumpsuit, leggings, romper, shorts, skirts, sweater, sweatpants, tank, tee, and top.

We achieved an accuracy of 74.25% in classifying garments to one of 18 categories. Typical errors were mixups between the same body part clothes e.g. tee and tank and this can be seen in more detail in the confusion matrix (Figure 2). If the 18 categories were grouped into three lower-level classes called “top”, “bottom” or ”long” based on the belonging body section, the classification accuracy increased to 97.95% as the task gets much easier. The accuracy estimates are most likely too pessimistic as we tested on images of varying filming conditions where the clothes were mostly on people, one image had sometimes multiple clothes partly visible, and there was some

[8] DeepFashion Database
http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html
[Accessed Aug. 21st, 2019]

[9] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang and Xiaoou Tang, "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1096-1104.
DOI: 10.1109/CVPR.2016.124

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. (2016) Deep Learning, Introduction, MIT Press.
https://www.deeplearningbook.org/

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot MultiBox Detector, Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham.

[12] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2261-2269.

amount of label noise in the data set (incorrect annotations). Some examples of our garment localization and classification results are shown in Figure 3.

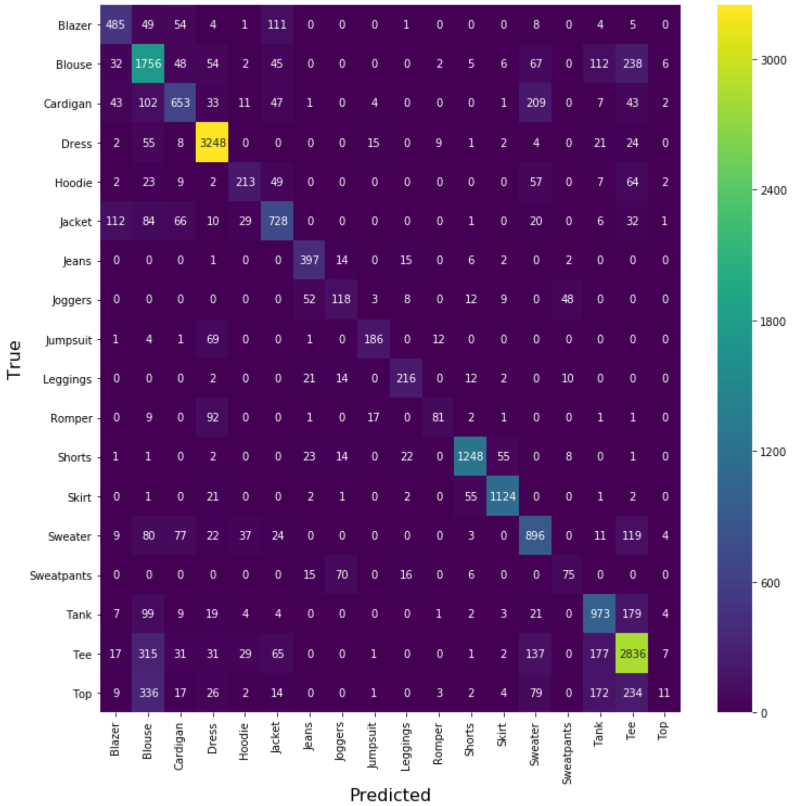

Figure 2. Confusion matrix presenting the frequencies of our classification predictions versus the ground truth garment class.

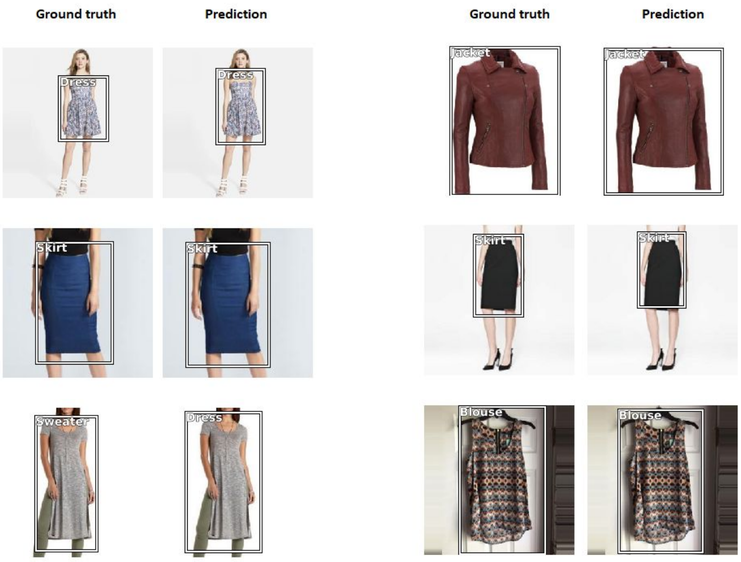

Figure 3. Garment classes with bounding boxes. Ground truth columns present the human annotations and prediction columns visualizes the prediction of our DenseNet-SSD model.

Although deep learning models are sometimes depicted as black boxes where we cannot understand how the model makes predictions, this is only partly true with CNN's. We visualized the neural network activations that contributed to the prediction decision on top of the image (Figure 4) with a method called Gradient-weighted class activation mapping (Grad-CAM) [13]. This confirmed our model had learned to held garment type-specific regions important. On a jacket, the model focuses on the collar and the armpit regions were especially salient with multiple examples. These make sense as collar gives important cues about the top-type garment class and the armpit region tells weather the clothing has sleeves. The fact that the model focuses on the garment and not on the person in images where the clothes are on a person, confirms that the model has generalized to detect clothes rather than body regions.

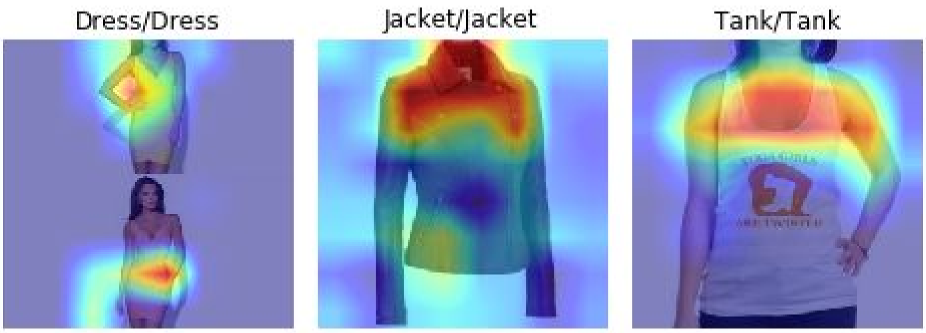

Figure 4. Grad-CAM visualization overlaid on the images. Title shows ground truth and prediction and colored regions highlights what areas mostly contribute to the prediction decision.

2.3 Measurement point detection

Measurement points are specific to the type of clothing. With a blouse, for example, we are interested to know the distance from left armpit to right armpit or from left hem to right hem. These aforementioned points do not apply to jeans or shorts, so we should have different measurement point sets for each garment type. It is necessary to recognize at least as many clothing categories as there are different measurement point set requirements. However, it could be beneficial to split some categories if their inter-class appearance varies very much. For example, hoodie and sweater probably have the same critical measurement points, but their appearance is very different when laid flat to a surface.

[13] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 618-626.
DOI: 10.1109/ICCV.2017.74

Many automatic garment measurement methods [1 2 3] detect corner or high curvature points from the garment contours and match these to a set of feature points in garment template by scaling and stretching the template to find the closest matches. This is problematic when a measurement point has low curvature in the contour, which can happen, for example, with shoulder points.

A task of locating predefined points in an image is called keypoint matching and most data science competitions of this type are currently won with CNN models [14 15]. The winning solutions of competitions are a good metric for techniques because compared to tens of published scores on known academic data sets yearly, these competitions can have thousands of data scientist teams participating from many different countries and focusing on the same problem.

The Vision & Beauty Team of the Alibaba Group and the Institute of Textile and Clothing of The Hong Kong Polytechnic University organized a garment key point detection competition in 2018 [15] and the winning teams achieved a normalized error of 3.3% which means that the average error of measurement point locations were around 3% of the width of the garment (armpit to armpit in top clothes and waistband to waistband in bottom clothes).

We trained a keypoint detector model of DenseNet121 base architecture where the model predicted the coordinates of six different measurement points: left & right collar, left & right hem, and left & right sleeve. We managed to get decent results (Figure 5) based on visual inspection and even locate points correctly inside the garment contours, that would not be possible using the contour point methods. Garment prints however, caused some error cases where a point was falsely detected from the edge of the print instead of a clothing edge.

[14] Facial keypoint detection challenge on Kaggle
https://www.kaggle.com/c/facial-keypoints-detection/overview
[Accessed Aug. 30th 2019]

[15] FashionAI keypoint detection of apparel
https://tianchi.aliyun.com/competition/entrance/231648/introduction
[Accessed Aug. 20th 2019]

European Union
European Regional Development Fund
Leverage from the EU 2014–2020

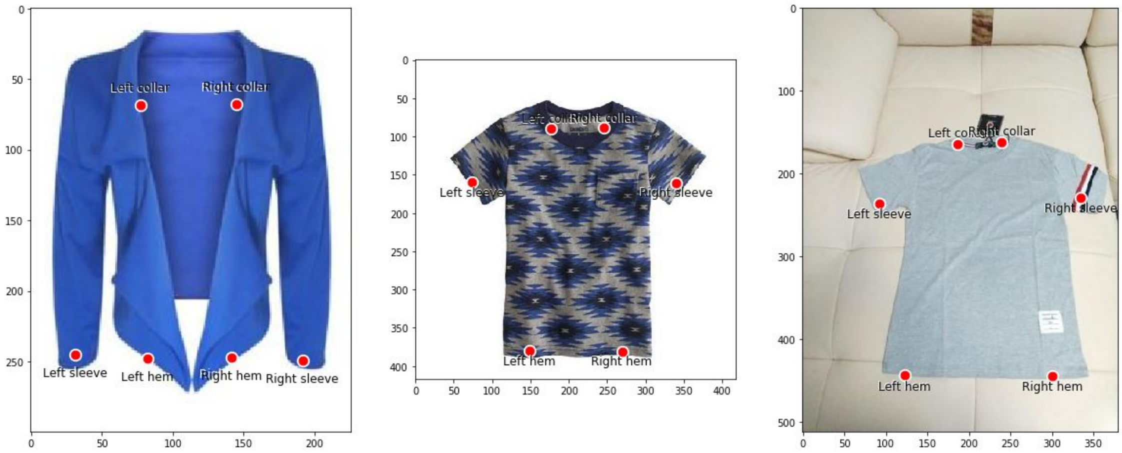

Figure 5. Predicted measurement point locations of our keypoint detection model. The annotated ground truth points are almost exactly the same as the predicted ones except for the rightmost t-shirts right sleeve point, which is incorrectly positioned.

We used the six measurement points for our model because the DeepFashion dataset [8] had annotations for them, although these points are not especially helpful for measuring clothes. In order to build a model for a specific set of measurement points, large amounts of annotated data are required. This means images of clothes where the pixel coordinates of the measurement points are specified, for example, with a CSV file linking the filenames with the coordinates. There are multiple keypoint annotation tools for images [16] [17] [18] and also annotation services [19] [20] that will provide annotations for uploaded images according to instructions.

3. Workbench specifications

The workbench serves a dual purpose in automatic clothing measurement. It provides a solid color background and good soft lighting for filming that helps to separate the clothes from the surface. The workbench is also necessary for reliable

[16] Visipedia annotation tool
https://github.com/visipedia/annotation_tools
[Accessed Sep. 9th 2019]

[17] COCO-annotator
https://github.com/jsbroks/coco-annotator
[Accessed Sep. 9th 2019]

[18] Supervisely
https://supervise.ly/
[Accessed Sep. 9th 2019]

[19] Mechanical turk
https://www.mturk.com/
[Accessed Sep. 9th 2019]

[20] Microwork.io
https://microwork.io/
[Accessed Sep. 9th 2019]

conversion from pixels to distance measure, and this is calculated from surface calibration marks or using a fixed camera to surface mounting with calibration shot.

If the camera mounting is fixed, the perspective transform matrix can be calculated only once, by filming a set of four marks of known object coordinates. Another option is to keep these marks, or other means of calibration, visible in each capture and calculate the transformation matrix every time. The latter does not expect a fixed camera mount so it can be used even with a handheld camera.

The camera's height and field of view (FOV) restricts the effective measurement area. The smaller the FOV, the higher the camera should be, so the lens should be carefully selected based on space restrictions. However, wide-angle lenses should be avoided as they increase the radial distortion, especially in the border regions. Also, keeping the camera's angle to the surface normal minimal reduces the possible distortion errors caused by rotational image transformation.

The surface of the workbench should be uniform in color and differ from the garment colors. Generally, a white background should do but even better if the surface color can be changed when necessary to provide maximum contrast between the clothes and the surface. Li et al. (2017) [2] used a LED lighted surface layer where the color and luminance of the surface could be adjusted by controlling the LED lights. The requirement of surface clothing separation is more significant when classical image processing methods are used for garment segmentation. Deep learning methods should learn to recognize the garment contours with less optimal conditions.

4. Error estimate

Possible error sources are either camera related, garment misclassification or measurement point localization related, or garment folding related. Camera related errors come from lens distortions or perspective transformation errors and are mostly determined by the success of calibration. The good aspect in camera errors is that they can be quantified by transforming the image plane calibration points back to object plane and calculating the difference. It is recommended to perform this type of sanity check after each calibration and specify a maximum error threshold that triggers an alarm for calibration error.

If the garment is misclassified to another category, the automatic clothing measurement can fail completely because relevant measurement points may not be retrieved at all. Measurement point localization related errors are measured as pixel offset or pixel offset scaled to a reference measure such as shirt or pants width (normalized error). With current state-of-the-art models, the expected normalized error is around 3.3% [15]. This means an error of 2.15cm with a sleeve length of 65cm and armpit to armpit reference measure of 45cm.

The final source of typical errors is garment folding. The image based measurement expects the garment to be flat on a surface and calculates the point distances from a surface aligned projection. However, not all garments fold perfectly flat by design or because of the thickness or filling of the material. Also, an image based measurement can only calculate the outer dimensions of the garments and these do not necessarily translate to the person's measures who wears the clothes.

5. Conclusions

Automatic camera based garment measurement is a feasible alternative to manual measuring. It offers a consistent way of taking the measurements because an algorithm measures from the same points regardless of who operates it. However, some amount of error can be expected due to various error sources and we estimate that 1 to 3 cm errors would be common depending on the length of the measure.

Building an automated garment measurement system that uses CNN models for category classification and key measurement point localization requires large amount of annotated image data. Large open datasets [8] can be used for initial training but the models should be fine tuned with an application specific dataset of at least 100 images per clothing category. This means using images taken with a workbench setup and measurement point locations manually annotated by experts.