Clothes Perception and Manipulation

D1.2
Demonstrator Procedures and Results, Annotated Data Specification

Version: 1.0
Last Update: 2014-12-31

Distribution Level: PU (Public)

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288553

Project Number	:	FP7-288553
Project Title	:	Clothes Perception and Manipulation

Deliverable Number	:	D1.2
Title of Deliverable	:	Demonstrator Procedures and Results, Annotated Data Specification
Nature of Deliverable	:	Report
Dissemination Level	:	Public
Version	:	1.0
Contractual Delivery Date	:	2014-12-31
Actual Delivery Date	:	2014-12-31
Contributing WP	:	WP1
Author(s)	:	Pavel Krsek (CVUT), Libor Wagner (CVUT), Vladimír Petrík (CVUT), Vladimír Smutný (CVUT), Tomáš Pajdla (CVUT), Václav Hlaváč (CVUT), Zbyněk Šika (CVUT), Petr Beneš (CVUT), Michael Valášek (CVUT), Jan Stria (CVUT), Christos Kampouris (CERTH), Andreas Doumanoglou (CERTH), Ioannis Mariolis (CERTH), Dimitra Triantafyllou (CERTH), Georgia Peleka (CERTH), Andreas Kargakos (CERTH), Vagelis Skartados (CERTH), Sotiris Malassiotis (CERTH), Gerardo Aragon-Camarasa (UG), J. Paul Siebert (UG), Li Sun (UG), Karel Křehnáč (NEO)
Reviewer(s)	:	Libor Špaček (CVUT), Sotiris Malassiotis (CERTH)

Abstract

This report describes the demonstration procedures of the CloPeMa project. The demonstration procedures consist of two joint scenarios and a few separate presentations. The two joint scenarios are motivated by laundry process. We introduce sorting of garments before washing and folding of a garment. The report contents list annotated datasets with short descriptions. The annotated datasets are results of the project.

Keywords

robot, manipulation, grasping, datasets, testing

Revision History

Version	Description	Author(s)	Date
0.1	Created	Pavel Krsek	2014-12-04
1.0	Proof Reading and Finalizing	Libor Spacek	2015-1-19

Contents

1 Introduction **1**

2 Final joint demonstrated scenarios **1**
2.1 Sorting the heap before washing 1
2.2 Folding the washed and dried garment 2

3 Other demonstrated procedures **3**
3.1 Compliant motion 4
3.2 Tying a knot 4
3.3 Hanging clothes over a string 4
3.4 Extracting a garment out of a stack of folded garments 5

4 Annotated data sets **6**
4.1 Colour and depth image dataset of spread garments 6
4.2 Garment folding photo dataset 6
4.3 Garment sorting dataset 7
4.4 Moving garment dataset 7
4.5 Glasgow's stereo dataset 8
4.6 Garment type and pose recognition datasets 9
4.7 Recognition and grasp point detection dataset for random forests 12

5 Conclusions **12**

1 Introduction

This Deliverable, number 1.2, describes the final demonstration procedures. These procedures allow us to demonstrate the ability of the CloPeMa test bed to manipulate soft materials. The procedures incorporate most of the algorithms and software packages developed under the CloPeMa project and represent results of the project. The procedures are represented by two joint scenarios which are described in section 2. The technical details of the realization of the demonstrator form part of Deliverable 7.4 [1].

Apart from the two main demonstration scenarios, we introduce a few additional separate presentations. These presentations show special developed skills 3, which could not be incorporated into the joint scenarios.

Annotated datasets were widely used during the development of various algorithms in this project. The datasets are one of the important public result of the CloPeMa project. We hope that the datasets will be used by other researchers. Therefore we prepared a description of each dataset and published the data on the net. The list of annotated datasets with their short descriptions is in section 4 below.

2 Final joint demonstrated scenarios

We would like to demonstrate the dexterity of the CloPeMa two hand manipulator. To this end we introduce two scenarios which demonstrate typical manipulation of a garment. According to the main idea of the CloPeMa project, the scenarios are motivated by laundry process. The first scenario represents sorting of garments, typically needed before washing. The second scenario is motivated by folding of a dry garment (after washing). Both scenarios are described in detail in the following sections.

2.1 Sorting the heap before washing

The sorting before washing scenario starts with a heap of garments. A creased garment is lying on the heap. The garment is picked up and its properties are measured. The garment is then classified and put on an appropriate (output) heap according to its type. The steps of the process are as follows:

Pick up a garment. Using Robot Head (UG): In order to detect grasping candidates on a heap of garments using the robot, we employ our surface topology-based feature extraction (described in D7.4 [1]). For completeness, we briefly describe our feature extraction technique in what follows. The feature extraction process includes four steps: B-Spline surface fitting, clothes topology and shape analysis, triplet matching, and grasping candidates ranking. As geometry-based features such as curvature and shape index are susceptible to high frequency noise, we firstly employ a B-Spline surface fitting approach to suppress noise induced from the cameras and the matcher. We then compute the shape index of a segmented clothing surface in order to detect its surface topographic features including ridges, contours, and surface shape types. Triplets are formed by matching ridge points with contour points, and grasping candidates are then ranked and selected. Figure 1 shows an example of the CloPeMa robot picking up a garment using the robot head.

Measure parameters. All sensors are employed to obtain all available parameters of the picked garment. We use force sensors on the wrist of the robot hand to measure the

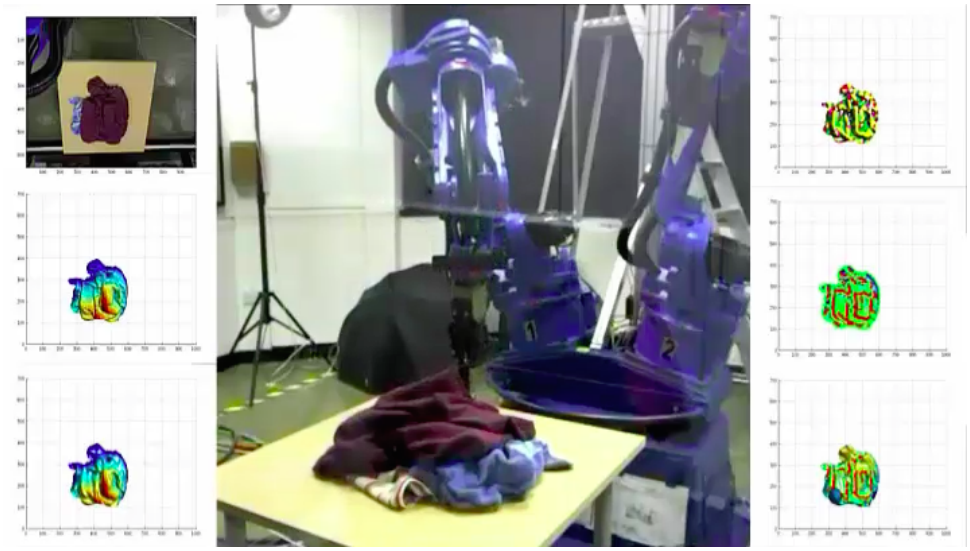


Figure 1: Picking up a garment using the robot head.

weight of the garment. Information about its colour is provided by camera. The photometric stereo sensor is used for obtaining detailed stereo image of fabric structure. Tactile sensor on the finger of CloPeMa gripper provides other parameters of the fabric structure.

Classify the garment. Based on obtained measurements, the garment will be classified. The classifier is selected based on desired sorting task. Operator can select the task from a list of basic sorting tasks. The tasks are: colour based sorting, sorting by fabric type or type of garment (T-shirt, trousers, ...). Combinations of these sorting tasks can also be used.

Put on appropriate heap. After the garment is classified, the manipulator puts the garment on an appropriate heap. Positions of the heaps are parameters of corresponding subroutines. The position can not be changed by the operator but can be easily redefined in the main level of the program code. The putting on a heap is realized by a movement which is planned each time with respect to actual collision model and heap position.

2.2 Folding the washed and dried garment

The folding process starts with a similar heap of garments as the sorting before washing scenario. The goal is a pile of folded garments. A garment is picked up from the heap by one hand. Then the garment must be unfolded and grasped in a well defined way. The garment is then put on a table and folded. The folded garment is finally put on a pile. Flattening is part of the process after the garment is put on the table. The steps of this scenario are:

Pick up a garment. Robot Head (UG): For this task, we employ the same robot skill described in 2.1.

Unfold the garment . First, the robot will hold the garment by one arm and grasp the lowest point by the other arm. Then it will rotate the garment in front of the depth camera, located on the belly of the robot. Captured depth images are used to recognise the type of the garment and the optimal position where to grasp it, using algorithm described in [3], in order to bring it into a predefined configuration. This is repeated until the predefined configuration of the garment is achieved.

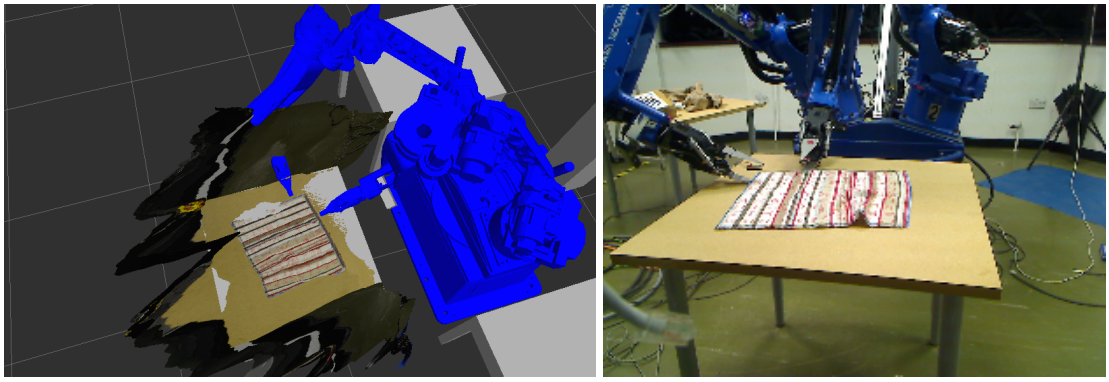


Figure 2: Flattening a garment using the robot head stereo.

Put on the table The robot will rotate from the unfolding area to the table. Then it will use the edge of the table to flatten the garment and place it on the table.

Flatten the garment. In order to flatten the garment, we employ the same feature extraction framework as that described in 2.1, with some subtle differences at the end. After a segmented clothing range map is B-Spline smoothed, we parse the garment topology by means of shape and topology into crease (or wrinkle) structures. Wrinkles length, width and height are used to quantify the topology of wrinkles and thereby rank them by size such that a greedy algorithm can identify the largest wrinkle present. A flattening plan optimised for this specific wrinkle is then formulated using dual-arm manipulation. Validation of the reported autonomous flattening behaviour has been undertaken and has demonstrated that dual-arm flattening requires significantly fewer manipulation iterations than single-arm flattening (as reported in D7.4 [1]). An example of this robotic behaviour is shown in Figure 2

Fold the garment The folding of a flattened garment is done in stages, one fold at a time. The number of folds varies with the type of the garment. At each stage the robot uses camera on a wrist to capture colour image of the table with the garment. The algorithm described in [5] is used to detect corners of the garment. Corners are used to compute the position of a fold. The robot will then execute the folding movement.

Put the folded garment on a pile First, the robot will rotate towards the table where the pile is to check the pile's position using Asus Xtion on its arm. This step is necessary as none of the cameras would see the table when the robot is holding the garment in both hands. Then rotates back towards the folding table, picks up the folded garment using the pick from side skill, rotates back towards the pile table and places the garment on top of the pile.

3 Other demonstrated procedures

We developed some unique robot skills which could not be sufficiently demonstrated in the complex joint scenarios or could not be incorporated into them. We decided to show these skills in separate demonstrations. These skills and demonstration procedures are described in this section.

3.1 Compliant motion

Compliant motion demonstrates the ability of the robot to be guided by a human operator. It uses the force sensor mounted between the gripper and the end of the robotic arm to measure the force acting on the gripper. Such a force is filtered, to remove the weight of the gripper and the noise, and then used to compute the desired velocity, as described in [2]. The computed velocity is then used to estimate the desired Cartesian position of the manipulator which is reached by linear interpolation. Should a collision or a kinematic configuration change be detected on the interpolated trajectory, then the motion is not executed to prevent unexpected behaviour. The demonstration example is shown in Fig. 3 and will be shown as an interactive demo.

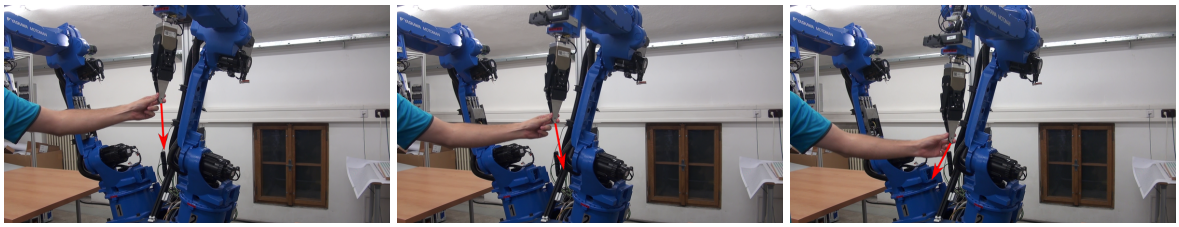


Figure 3: Compliant motion sequence

3.2 Tying a knot

In this demonstration we show the robot tying a knot on a rope. At first both ends of a rope are placed into robot's grippers. By following fixed trajectory, the rope is placed on one of the robot's wrists (see Figure 4) so that it creates a loop. One end of the rope is released so that it falls into the middle of the loop. The colour and depth image from Asus Xtion on a wrist of the free arm is used to find the free end of the rope. Then the free end is grasped through the loop by the free arm. The hitherto unmoving arm is moved into a position that will allow the rope loop to slide from the wrist thus finishing the not. To tighten the knot the robot spreads its arms until a force threshold is exceeded.

3.3 Hanging clothes over a string

The goal of this skill is to hang clothes over a string (eg. a washing line). Special holder was made for this purpose. It is constructed from two legs, stable base and a string drawn at the top between the legs. ROS visual and collision model was created for this holder and the string. It is shown in Figure 5.

The length of clothes which are going to be put over the string is found using 3D data from Xtion which is mounted on the robot torso; it sees any clothes in front of the robot.

The detected length is then used to calculate the final position of the grippers before hanging the clothes over the string. It is also used to estimate the weight of the clothes and its influence on the string bending, when the weight is not previously known.

An example result of hanging clothes over the string can be seen in Figure 5.

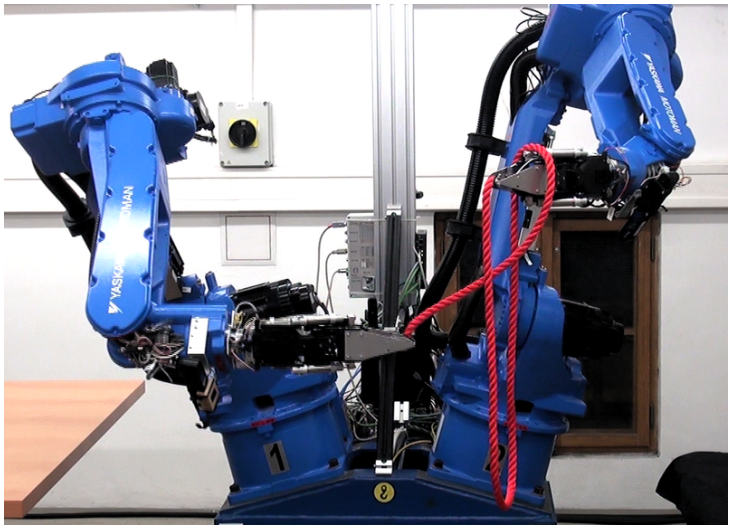
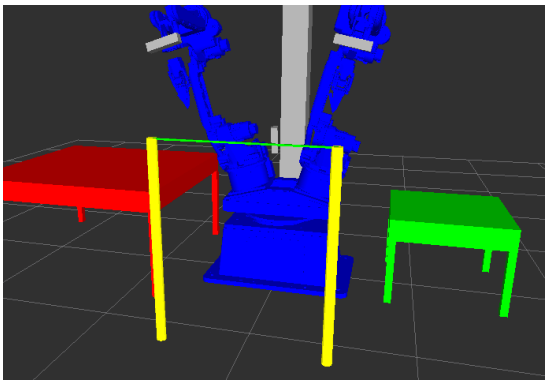
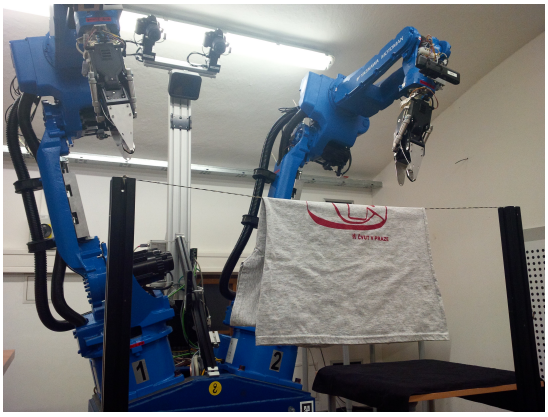


Figure 4: Tying a knot



(a)



(b)

Figure 5: Hanging on a string (a) ROS model, (b) clothes on a string

3.4 Extracting a garment out of a stack of folded garments

The goal of this skill is to pull a specific garment out of a stack. The initial state is a stack of folded garments placed on a table. The first step is *Planar Extraction* using [4], to remove the points representing the table, the floor and the wall from the initial point cloud. The remaining points represent the stack of garments. We apply to them *Color Segmentation* to separate each garment as a unique cluster of points using the *Color-Based Segmentation of Point Clouds* algorithm [6]. Each garment is located as a unique cluster. Two grasping candidates for each garment are identified. The grasping candidates are computed for each cluster using the following two assumptions: (1) the grasping points are extreme points in the cluster, (2) they are located in the area between two clusters representing consecutive garments.

An example of removing the third garment out of the stack is shown in Figure 6, where the grasping candidates are shown as vectors. Using these grasping candidates, the robot removes all garments placed above the preferred one and places them as a second stack nearby on the table. The preferred piece is then extracted and the remaining unselected pieces are placed on top of the remaining garments on the initial stack.

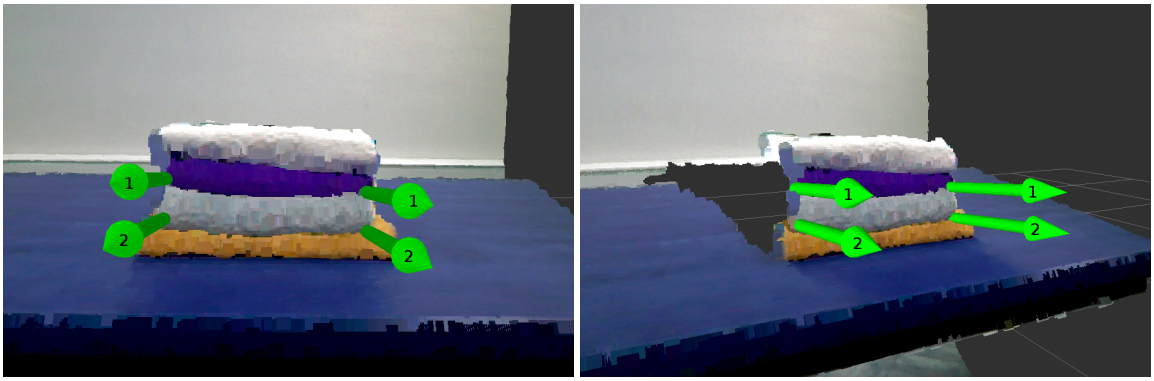


Figure 6: Illustration of grasping candidates for removing the third garment out of the stack. Candidates are shown as perpendicular vectors to Asus Xtion's frame. Using these vectors the robot arms rotate to the sides of the stack, perpendicular to the vectors. First they extract the top two garments using vectors labelled as "1", and then the selected garment, in this case the third one, is extracted using the vectors labelled as "2".

4 Annotated data sets

Annotated data sets are the foundation of the implementation of some of our developed algorithms. We collected different types of data sets during the CloPeMa project. In our opinion, the data sets are unique and could be interesting to other researchers. We feel that the annotated data sets are important public results of the CloPeMa project, which can be used even a long time after the project finished.

Therefore, we are describing the data sets and publishing them on the net. Most of the annotated datasets mentioned in this section have been published already. References to the public data sets are in the following text. We expect that the rest of the data sets will be published with our works which are based on the data.

4.1 Colour and depth image dataset of spread garments

This publicly[1] available dataset consists of images of spread and folded garments. The images are taken with Asus Xtion placed on a robot arm. We have used several types of garments including shirts, t-shirts, trousers, skirt, towel and others. The garments are captured in several configurations, spread and folded. Each image is manually annotated with garment model including position of named corners, for folded images also position of folds.

The dataset is intended to train and test algorithms for garment segmentation, recognition, model fitting and garment folding.

4.2 Garment folding photo dataset

A publicly[2] available dataset of colour and depth images from various stages of garment folding. Colour and depth images were taken after each folding step using both Asus Xtion and Stereo Head camera. The dataset consists of 121 folding experiments with towel and t-shirt. Each experiment was manually annotated with a success tag and a list of imperfections of the folded garment.

[1] `http://clopema.felk.cvut.cz/color_and_depth_dataset.html`
[2] `http://clopema.felk.cvut.cz/garment_folding_photo_dataset.html`

The main purpose of this dataset is to record failures in the folding process. This information is then used to implement robust algorithms that will prevent or recover from these failures.

4.3 Garment sorting dataset

This publicly[3] available dataset consists of various sensory outputs, such as colour image, weight, tactile, detailed stereo image and depth image of a garments. Each garment was picked by the robot, then several sensors were used to capture information that may be used to sort garments by several factors.

The dataset can be used to train machine learning algorithms for garment type and material estimation. This information is then to be used to sort garments for laundry.

4.4 Moving garment dataset

For the purpose of the estimation of dynamic parameters of a cloth, we have captured several videos of moving garments by depth camera. We have used two configurations, in the first configuration the garment was moving parallel to the optical axis i.e. towards the camera, in the second configuration the garment was moving perpendicularly to the optical axis i.e. horizontally in the image plane.

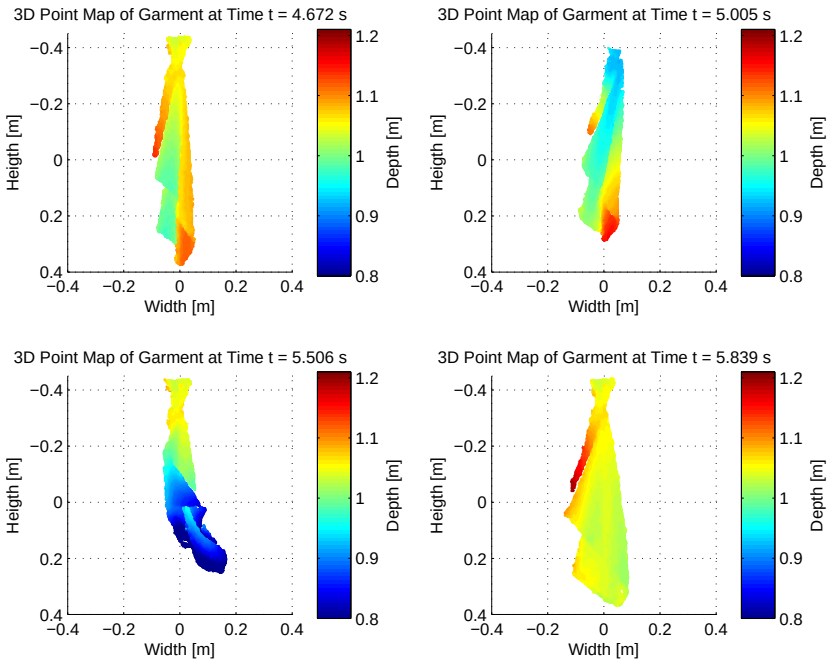


Figure 7: Moving garment dataset

The dataset consists of 33 ROS bag files that were recorded through the experiment. Each bag file contains periodic information about the robot joints state, all robot frames position, depth images. Various garments we used, including towel, t-shirt, shorts, skirt and pullover. Examples of the captured data are shown in Figure 7.

[3] `http://clopema.felk.cvut.cz/garment_sorting_dataset.html`

The dataset is not yet publicly available but we are prepared to make it public along with the submission of related papers.

4.5 Glasgow's stereo dataset

This publicly available dataset[4] consists of 80 stereo-pair colour images of garments over different pose configurations with their corresponding horizontal and vertical disparity maps and mask annotations. This database is based on 16 off-the-shelf garments. Each garment has been imaged in five different pose configurations using the CloPeMa binocular robot head. Figure 8 shows a subset of the dataset. The aim of this database is to serve as a benchmark tool for algorithms for recognition, segmentation and various range image properties of non-rigid objects. The dataset is free for research and educational purposes and can be used in scientific publications on the condition of respecting the requested citation acknowledgement.

Figure 8: Glasgow's stereo dataset.

[4]https://sites.google.com/site/ugstereodatabase/

4.6 Garment type and pose recognition datasets

Two large datasets of RGB and depth images of hanging garments have been created. The first one contains images of actual garments hanged by a robotic manipulator, whereas the second dataset contains syntetic images of garments simulated using 3D computer graphics software. These datasets are publicly available for download[5]. A short description of each dataset is provided bellow.

Real Dataset This dataset contains 154620 RGB images of hanging garments in ".png" format and their corresponding depth images (also in "png" format) acquired by an ASUS XtionPro range sensor. In order to acquire these images, a robotic manipulator has grasped the garments and held them in front of the XtionPro sensor. Before grasping, markers have been placed on the garments denoting different poses on the unfolded configuration. Then, using the manipulator, the hanging garment has been rotated 360°, while the XtionPro sensor acquired 180 RGB and depth images. Image size is 640x480 pixels for both modalities. Robot grasping has been performed manually, with the robot holding the garment from every marker.

Acquisition has been performed for 19 different garments belonging to 5 garment types. More specifically, we employed 3 pairs of trousers, 4 shorts, 3 shirts, 5 t-shirts, and 4 towels of different sizes and proportions. We have defined 59 poses for trousers, 20 poses for shorts, 74 poses for shirts, 56 poses for t-shirts, and 25 poses for towels.

An example of the resulted configurations and corresponding first poses is presented in Figure 9 right. For each garment an RGB image depicting its unfolded configuration and the markers' topology is included, while a list of coordinates of the markers in the image is provided in a '.txt' file. The scale ratio between pixels and cm (10mm) is also provided in the end of the ".txt" file for easy conversion between them.

A separate directory is created for each garment type (named after the type). It contains two subdirectories named 'rgb' and 'depth', where RGB and depth images of the garments are stored, respectively. In each of these subdirectories the corresponding ".png" files are named using the following notation: GN_GT_PN_VN.png, where GN denotes the serial number of the garment, GT denotes its type, PN denotes the pose number, and VN denotes the view number. Thus, the first depth image of a towel is stored in the database in

/real_garments/towel/depth/03_towel_01_001.png.

Type recognition labels should be constructed using the GT values, whereas pose recognition labels should be constructed using PN values. In order to segment the garment from the background, depth images should be thresholded keeping only values between 700 and 1400 mm.

Synthetic Dataset. We have constructed a large dataset (SD) of synthetic depth images using Blender 2.6.2, an open-source 3D computer graphics software. We have constructed 48 models of shirts, 48 models of trousers, and 48 models of towels. The models of the same category differ in shape, mass, size and material properties. In order to simplify and speed-up the simulation process we are using 2-D models of the garments, assuming that the front and back sides of the clothes are not separated. Even with this simplification, the models approximate surprisingly well the configuration of real garments hanged under gravity (see Figure 9). An example model for each category along with their corresponding triangular mesh is presented in Figure 10. The mesh of the shirt models consists of 141 vertices, whereas there are 113 vertices for pants and 81 vertices for towels. Due to symmetry the above vertices correspond to 74, 59 and 25 poses respectively.

[5] http://clopema.iti.gr/datasets/DeepGarmentRecognition

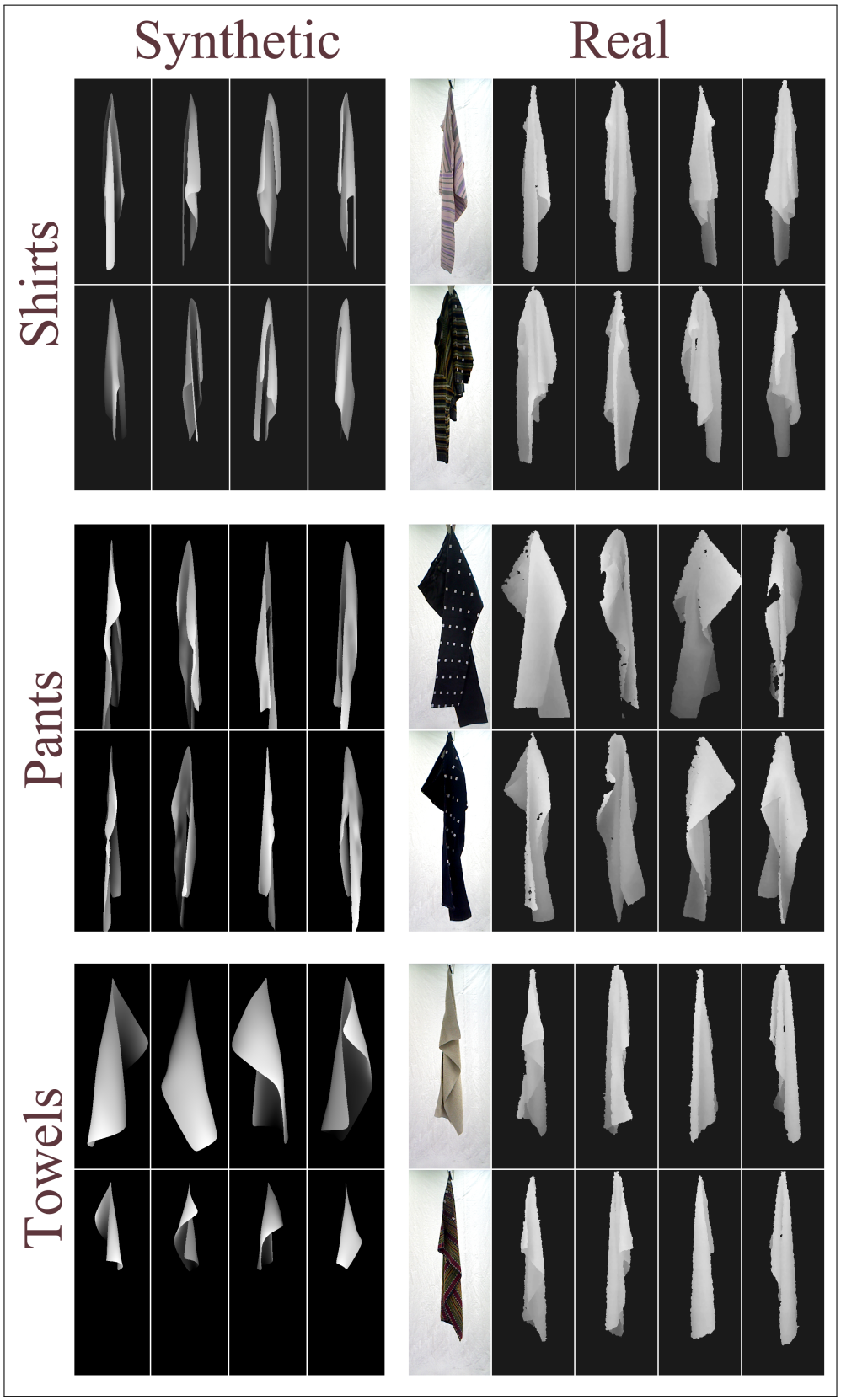


Figure 9: Example images of synthetic (left) and real (right) garments hanged by their first pose.

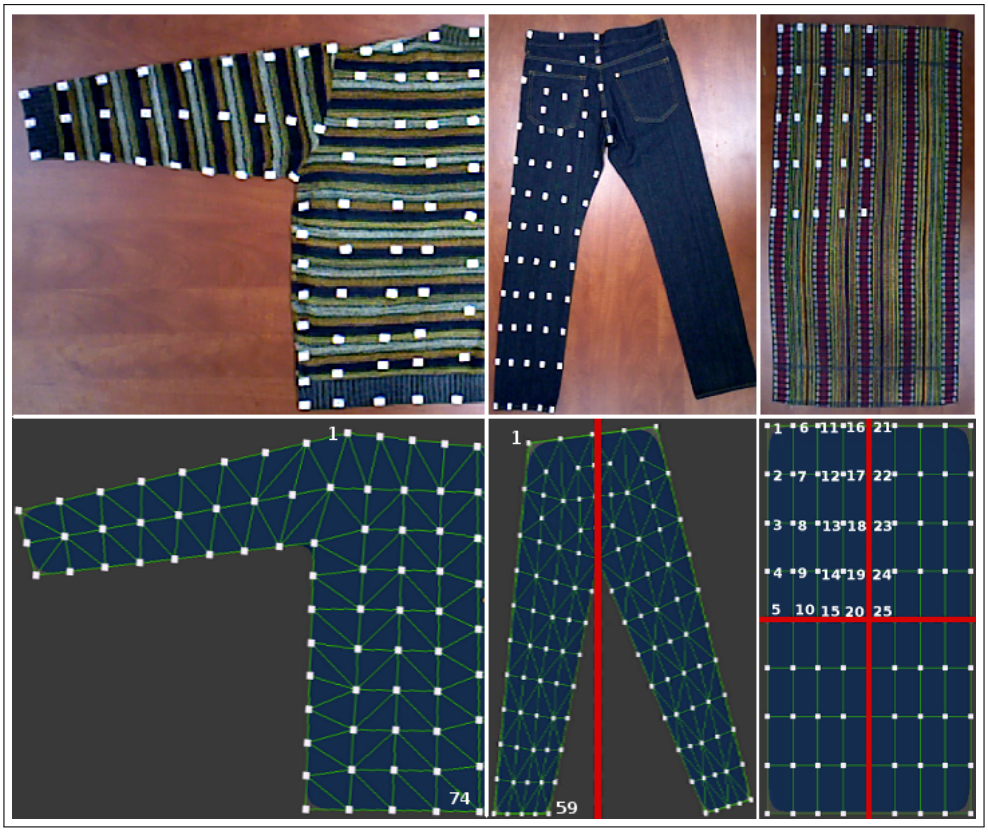


Figure 10: Selected vertices defining the poses of the hanging garments. In the top row, white markers define the poses on real garments. In the bottom row, the synthetic models are depicted. The selected poses are defined by their vertices that reside to the left of their symmetry axes.

Using Blender's cloth engine, we have simulated hanging the models by each of their vertices. A ring of 80 virtual cameras is placed uniformly around the hanged garment in such a height that they can view the entire garment, and 80 synthetic depth images are acquired. The use of the camera ring removes the need for rotating the garment, which introduces additional noise during acquisition with the real system that uses only a single sensor. The formulation of the synthetic dataset needs only a small fraction of the time needed by the real world system, whereas the resulted images are noise-free. The set of the 80 images for a single vertex is acquired in about 1 min. Namely, the entire database of the 1286400 depth images (in ".png" format) has been created in about 12 days. The virtual cameras were positioned at a distance of about 1060 mm from the hanging axis, whereas their vertical distance from the hanging point was set to 400 mm, equal to one of the robot's XtionPro sensor. The virtual camera's intrinsic parameters where also matched to those of the XtionPro sensor.

An example of the resulted synthetic configurations and corresponding first poses is presented in Figure 9 left. For each model with different shape, an RGB image depicting its unfolded configuration and the vertices' topology is included, whereas a list of coordinates of the vertices and the selected poses is provided in a ".txt" file.

A separate directory, named after the type, is created for each garment type. It contains six subdirectories named TYPE_1, TYPE_2, etc., corresponding to the six different shapes selected for the models of this garment type. In each of these subdirectories the corresponding ".png" files are named using the following notation: GT_ShI_MI_BI_SI_PN_VN.png, where GT denotes its type, ShI denotes the index number of the model's shape, MI denotes the index

number of the model's mass, BI denotes the index number of the model's bending parameter, SI denotes the index number of the model's structure parameter, PN denotes the pose number, and VN denotes the view number. Thus, the first depth image of a towel model is stored in the database in

/synthetic_garments/towel/towel_1/towel_1_1_1_1_1_1.png.

Type recognition labels should be constructed using the GT values, whereas in pose recognition labels should be constructed using PN values. In order to segment the garment from the background, depth images should be thresholded keeping only values between 700 and 1400 mm.

4.7 Recognition and grasp point detection dataset for random forests

This dataset contains images of clothes hanging from all possible lowest points. It is used for testing the unfolding procedure based on Random Forests for the final review. It contains 4 different garments: two shirts, one pair of shorts and one towel. The folder names correspond to each garment hanging from the lowest point, and when the folder name has "2" in the end, the garment is grasped from the 1st desired point. Each folder contains one folder called "depth" and one file called "*garment*_annotation.txt". The images in the folders are in PNG format and are named as "tryNo_viewpointNo.png". The variable *tryNo* counts the times that the robot picked the lowest grasp point, and *viewpointNo* counts the different viewpoints that the garment is captured from, while being rotated by the gripper. The annotation file contains several lines of the following format: *tryNo viewpointNo lowestpointNo* X_{grasp} Y_{grasp}. The variable *lowestpointNo* counts the different possible lowest points, and point (X,Y) corresponds to the location on the image of the desired grasp point for unfolding. In case it is hidden the values are set to -1. The total number of images is 10995.

5 Conclusions

This report presents two joint demonstration scenarios which documents the capabilities of the CloPeMa test bed. The scenarios are motivated by garment manipulation during a laundry process. Some other test bed developed skills are shown in separate demonstrations, which were described.

The list of annotated data sets is an important part of this report. The annotated data sets were collected during the CloPeMa project and are public results of the project. Most of the data sets are published now. The rest of the data sets will be published with their corresponding articles. We expect that the public data sets will be interesting for other researchers.

References

[1] Deliverable D7.4 - Third project demonstrator and its validation, European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 288553, CloPeMa, December 2014.

[2] Deliverable D2.3 - Motion planning based on advanced sensors and reasoning, European Communitys Seventh Framework Programme (FP7/2007-2013) under grant agreement no 288553, CloPeMa, January 2015.

[3] Tae-Kyun Kim Sotiris Malassiotis Andreas Doumanoglou, Andreas Kargakos. Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 1703–1708. IEEE, 2014.

[4] Dirk Holz, Stefan Holzer, Radu Bogdan Rusu, and Sven Behnke. Real-time plane segmentation using rgb-d cameras. In *RoboCup 2011: Robot Soccer World Cup XV*, pages 306–317. Springer Berlin Heidelberg, 2012.

[5] Jan Stria, Daniel Průša, Václav Hlaváč, Libor Wagner, Vladimír Petrík, Pavel Krsek, and Vladimír Smutný. Garment perception and its folding using a dual-arm robot. In *Proc. International Conference on Intelligent Robots and Systems (IROS)*, pages 61–67. IEEE, 9 2014.

[6] Qingming Zhana, Yubin Liangb, and Yinghui Xiaoa. Color-based segmentation of point clouds. *Vol. XXXVIII, Part 3/W*, 2009.