World representation for a dual-arm robot manipulating with clothes

Jan Šindler, Václav Hlaváč
sindljan@fel.cvut.cz; hlavac@fel.cvut.cz

CTU–CMP–2013–05
January 20, 2013


This research was supported by the European Commission FP7 project ICT 288553 CloPeMa (Clothes Perception and Manipulation).

Research Reports of CMP, Czech Technical University in Prague, No. 5, 2013

Published by Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University Technická 2, 166 27 Prague 6, Czech Republic fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz
Abstract

The work deals with the task of robotic perception/manipulation with soft materials, in particular with clothes and garments. It is part of the diploma thesis project of the first author. The report serves two main purposes. First, it surveys related work. Second, it outlines a subproblem, namely the task how the cloth (garment) should be modeled/represented to enable robotic tasks such as folding the garment, sorting a heap of washes coming out of a dryer. The full scale physical modeling (based on partial differential equations and their finite elements method solutions) is given up because of its complexity.

Our approach is based on the assumption that the cloth/garment has natural and relatively stable intermediate states of minimal energy. There are basic action which allow to transfer the cloth from such a stable state to another one. It is also believed that set o basic actions performing such transitions can be designed.

The actual proposal of the report is in outlining possible representation of the manipulated cloth.
Contents

1 Motivation 3

2 Goals 3

3 Cloth representation 4

4 Scene representation, reasoning engine 7
   4.1 External goal and initial parameters 7
   4.2 Collection of pre-learned actions 8
   4.3 Function storage 8
   4.4 Reasoning engine 8
   4.5 Robotic middleware, ROS 8
   4.6 Expected sensory inputs 9
   4.7 Output of the system 9

5 Work of others 9
   5.1 A global point of view 9
   5.2 Relevant projects 10
   5.3 Cloth modeling 10
   5.4 Hardware scope 11
      5.4.1 Used manipulators 11
      5.4.2 Used grippers 11
      5.4.3 Used sensors 12
   5.5 How the others manipulate cloth 12

6 Conclusions 15
1 Motivation

The main motivation for this work came from interesting research tasks formulated in the European project CloPeMa [2]. The project is focused on garments manipulation by a two armed industrial manipulator. This diploma project is supposed to propose a representation of a robot world model, which will help perception/manipulation tasks. The ability of a system to self-learn from gained experience is expected too. The proposed model should be independent of used sensors and a particular robot. The model independence will be tested on two robots, first at my home Czech Technical University in Prague and second at KTH in Stockholm during my several month long Erasmus stay starting in a second half of January 2013. A functionality of the proposed robot world model will be demonstrated on an example of a towel folding, which was performed by others before.

2 Goals

CloPeMa project finds new methods allowing easier perception/manipulation with clothes, textile, garments. The goals are as follows:

1. Create a robotic device that will be able to manipulate with garments.
2. Select proper types and amounts of sensors.
3. Select a proper type of a gripper.
4. Propose and implement a cloth model that can be used for a higher level of manipulation planning.

Goals of the diploma project reflect a part of CloPeMa goals. Several institutions such as Centre for Research and Technology Hellas, Informatics and Telematics Institute in Thessaloniki, the Czech Technical University in Prague, the University of Glasgow, University of Genova cooperate on CloPeMa project.

We would like to contribute to the cloth model task and model verification. My goals are as following:

1. Propose and implement the model of a cloth that can be used for towels and alike.
2. Implement algorithms that will help to create the model. Examples of useful algorithms are:
3 CLOTH REPRESENTATION

- Cloth corner detection which was proposed in Shepard [18].
- Hem detection as in Hamajima [16] and other useful algorithms.

3. Detect grasp points in the model.

4. Based on this information, decide where the towel can be grasped.

5. The proposed system should be tested on two different robots with various sensors.

3 Cloth representation

A cloth or garments made from a cloth are non-rigid objects. A stability of its’ shape cannot be guaranteed while moving or touching it. However, there are relatively stable cloth configurations matching to a ‘low energy state’. Consider an example of a piece of cloth lying on a table or hanging on a hanger.

Use of these stable configuration is the proposed way how to untangle the inherent complexity of a cloth physical model. All of the models mentioned in Section 5.3 are too complicated and slow for a real time computation or not general enough.

Figure 1 depicts a subset of clothes used in CloPeMa project. The expectation is that the model will be able to represent the scene and objects from this class.

![Figure 1: An example of a subset of clothes used in CloPeMa project.](image)

This diploma project and also CloPeMa project seek some shortcut representation, which would help us to model cloth. The representation/model should be less computationally intensive than the physical modeling. This is our motivation to propose an alternative cloth representation.
Stable configurations of a piece of cloth provide relatively well defined situations, which can be used to simplify both modeling and acting on a model (e.g., manipulating a piece of cloth). Such stable configurations allow simplifying the cloth perception/manipulation tasks into well defined subtasks, for which basic actions can be prepared/learned in advance. Such basic actions can be, e.g., grasping by a single or two hands in grasping points, lifting the cloth to obtain another stable configuration, disarranging a heap of garments if it is difficult to perceive individual pieces in it.

We suggest to split a cloth model into four representation layers ordered by the degree of semantic information needed to solve the cloth perception task at hand. These layers are processed in a a bottom up manner. The structure as shown in Figure 2.

The first layer stores raw sensory data. We call it a “raw data storage”. The first layer also keeps the following information:

- What is the source of the data.
- When the data was measured.
- What was the configuration of a sensor during the measurement.

The second representation layer is the “processed observation”. The system stores data computed from raw data or another inherited data by some of the available algorithms. For example, the system finds a manifold representing the visible part of the garment from the input point cloud provided by a stereo vision system or the rangefinder (e.g. Kinect). The benefit of the manifold is that the created surface is coherent and, if a proper algorithm was chosen, does not contain holes. The use of manifolds and a different way how to create them was described in Berger et al. [12]. The system also stores the following information in processed observations layer:
• A pointer to the source of data.
• A list of functions used to compute these data.
• Time stamp when the observations were computed.

The “cloth tokens” constitutes the third layer. The cloth tokens generalize the important data features from its use point of view. The example of tokens are grasping points, detected features such as sleeves, hems. If we consider a towel example, it will contain the position of hems and position of detected cloth corners.

The upper layer is the “topological map” represented by a graph. Two hypothetical examples of the topological maps are shown in Figure 3. The topological map represents a physical bindings between “cloth tokens”. The left part of Figure 3 depicts one possible topological map, in which connectedness (topology) is given by the semantics embedded to the outline of a single region.

![Figure 3: An example of two possible topological maps of a shirt with long sleeves.](image)

The right part of Figure 3 demonstrates that the image can be segmented into several regions which have clear semantics. A more ‘semantic’ topological binding is considered here. It is expressed in Figure 3 by blue ‘spectacles’ expressing, which labeled regions are semantically connected each to other. The topological map represents the state of scene part, which the robot needs to understand, and potentially manipulate objects in it.

Of course, it will be much more difficult to interpret a realistic scene with garments, e.g. a heap of garments taken from a dryer, in terms of topological maps. It will be likely impossible just from observing the scene. However, a smart manipulation with objects in the heap is the way to untangle this puzzle.
4 Scene representation, reasoning engine

Let us consider the reasoning engine to be a black box for a simplicity now. Let assume that the reasoning engine is able to interact with the proposed world model and perform some reasoning on it.

The cloth model, which was already introduced in Section 3 and the model of the nearest robot environment is of my concern. Both models are used in reasoning process. The data are read from it and written to it too.

The robot system has those or that sensors and actuators allowing to sense/modify its world. The example of the current imperfect CloPeMa robot actuator is depicted in Figure 4. The reasoning engine, sensors and actuators maintain the equivalence between the reality and the robot internal model of it.

Figure 4: The current imperfect actuator of CloPeMa robot, basically a pair of tweezers.

4.1 External goal and initial parameters

It is assumed that the overall goal (task to be performed) comes from the outside of the reasoning engine, probably from a human user. The following parameters from the outside initialize the system:
• The goal, provided either by a person or, potentially, by a higher level reasoning module.
  
  The example of a goal relevant to my work is “fold the towel”. The user has to tell the system how does the folded towel look like.

• What kind of actuators the system has. Particular actuators imply the admissible actions.

• What sensors and/or sensory information is available to the system.

4.2 Collection of pre-learned actions

The space created by all possible cloth/garment states and actions, which a robotic system can perform, is too large, difficult to model, and to plan.

My strategy is to bypass the complexity by decomposing the complex problem into several simpler subproblems. The transitions between subproblems traverses through stable states, e.g. the energy minimizing configurations of the cloth or garment. Transitions among these stable states induce a set of pre-learned basic actions, for example pick-up a piece of cloth from a table, grasp a sleeve or trouser-leg with one fold and unfold it, and so on.

The long term aim could be to establish the set of basic actions automatically. However, this would be too ambitious goal for this work. Having in mind that this work concentrates to the representation of cloth/garment, We will prepare several basic actions manually, probably by using a teach-pendant (a hand-held robot control terminal) of the robot.

4.3 Function storage

The system stores functions that converts data from one form to some another. A list of functions, for example a function that converts the point-cloud (output of a Kinect, for example) into the grasping points. Knowledge about functions is available to the reasoning engine.

4.4 Reasoning engine

The reasoning engine will read/write data from internal representation, sensors, actuators and perform action planning.

4.5 Robotic middleware, ROS

There are good practical reasons for using a middleware, which abstracts the robot and allows to interconnect different control and sensory modules
4.6 Expected sensory inputs

Incoming sensory information updates continuously the world model including observed objects in it. The expected inputs are:

- A picture/video from a camera or several cameras;
- A point cloud from Kinect(s) and from a stereo-vision camera head;
- Information from tactile sensors.

Besides raw sensory data, the higher level module will likely know a wider context of a particular sensory information. As a minimal requirement, the higher level module will keep and update the information about the format, in which the particular sensor provides its observations.

4.7 Output of the system

The expected output should be the sequence of actions. Let consider a towel as an example. The towel lays on the table in an undefined form at the beginning. The system should provide a sequence of actions, which lead to a folded state. The expected partial output would be an initialized virtual model and information that enriches our database used for object recognition and decision making.

5 Work of others

The topic we work on can be looked at from several views. The description of the views is divided into the following sections. Section 5.1 gives to us information about existing surveys that are connected with our work. Section 5.2 lists projects that are similar to ours. Section 5.4 is focused to used hardware such as manipulators, grippers and sensors. Section 5.3 describes existing cloth models. Section 5.5 summarizes techniques, which other researchers used for a cloth detection and manipulation.

5.1 A global point of view

A summary about what was done in a dual arm manipulation can be found in Smith et al. “Dual arm manipulation survey” [21]. Hu et al. “Review of
Cloth Modeling" [?] contains useful links about a cloth 3D approximation from patterns, a cloth approximation by its profile and a parameterized model of cloth.

5.2 Relevant projects

There is a research group at the University of California, Berkeley, Department of Electrical Engineering and Computer Science which has become famous for towel folding videos since 2010. Relevant articles are Shepard et al. [18] and Towner et al. [14]. More information, some datasets and additional articles can be found on web pages [8]. Web pages [7] provide also some information about cloth modeling and manipulating.

- STIFF [1] - Its aim is to understand and mimic a variable stiffness paradigm that is used by a human nervous system. The goal is to equip an agile, robust and versatile robot hand-arm.

- VIACTORS [10] - This project aims at developing and exploiting actuators for a new generation of robots that would be able to co-exist and co-operate with people.

- LEAPFROG [6] - A module B of this project named “Automated Garment Assembly” is closely relevant to CloPeMa. It was focused on better handling of limp materials (fabric).


- DEXMART [3] - The project is focused to areas where dexterous and autonomous dual-hand manipulation are required.

Articles written about clothes manipulation are rather frequent and can be found in Section 5.5.

5.3 Cloth modeling

We found only a little information about cloth modeling. A point cloud cloth model has been used in [18] and [13]. “Graph of connected hems” have been used in [16] as a cloth model. A lot of models based on cloth physical parameters such as elasticity were used in 3D screen rendering. Examples of its' usage can be seen in the Au [?] or the Chen [?]. Almost every work in this area uses Flexible-Distortion model, Particle-System model, Finite Elements method, Spring-Mass model or alike.
The cloth description/model, which is closer to our idea, was introduced in Miller [22]. Their model is based on a parameterized shape description of a cloth. The authors used landmarks detected on the piece of garment and determine constraints between a position of it. They proposed a separate model for each garment type/class (shirt, pants, ...). The model contains a minimal set of landmarks, each augmented with its position. These landmarks should distinguish between the classes. Wang [24] published an extension of the model, which added local features (such as cloth texture) into it. A more detailed description of both approaches can be found in Section 5.5.

5.4 Hardware scope

Almost everybody used two armed static or mobile manipulator with 6 or 7 DoFs. The static platform has been used in this work [20]. The mobile platform exists in two forms a wheeled form [18], [23], [14] and a humanoid form [15].

5.4.1 Used manipulators

A PR2 robot [11] or some modification of it was the most favorite device used for experiments with garments. The PR2 robot has been used in works [18], [23], [24], [22] and [14]. Kawasaki JS2 and Yamaha Rch-44 were applied in the Salleh [20] work. Asimo [5] in the Gienger [15] work was the only one other robot we have found.

5.4.2 Used grippers

Almost every researchers, the research topics of which is not connected with a “How a garment should be held” use standard type of gripper with:

- 2 rotation fingers [13], [18], [14], [24], [22];
- 3 rotation fingers [19];
- 2 translation fingers [25];
- Articulated hand [15];

Two types of the grippers, which were not so usual and were designed specially for cloth manipulation were found. These are the Inchworm gripper [20] and the wheeled gripper [16]. The inchworm gripper is able to move along a garment hem easily. The wheeled gripper can pick up a cloth easily, which lies flat on a table and furthermore it does not need so precise detection of a grasping points.
5.4.3 Used sensors

A very wide range of used sensors can be found:

- A Kinect or something that works on the same principle is described in Ramisa [19] - They used Kinect for a depth map acquisition.
- A stereo-vision camera head [18] - It is also used for a depth map acquisition.
- A stereo-vision with a pattern projection [17] - This works on the same principal as Kinect but not in the compact body.
- A single (vision) camera is used as an image source for corner detection in Salleh [20]. The camera was also used as input source of data for features classification and decision making in Wang [24].
- A pressure sensor [13] - Pressure sensors in this example are used for detection that cloth has been grasped correctly.
- An infrared sensors [20] - The IR sensors are used for a cloth corner detection inside a gripper.

5.5 How the others manipulate cloth

We list approaches how others dealt with the cloth perception and manipulation process.

- Maitin-Shepard [18], which has got a lot of popularity in press, a new technique for a cloth corner detection was proposed. The input were images from a stereo-vision camera head. Depth-discontinuity points and a depth-discontinuity map were computed from the input images. Authors used a sharpness of a curvature to distinguish between folds and edges of a cloth. The process they have used was: picking up a cloth from a pile $\rightarrow$ shaking the cloth to put it into the low-energy configuration $\rightarrow$ putting cloth into the basic configuration (corner detection, re-grasping and twisting) $\rightarrow$ folding towel in an open loop sequence of previously defined moves.

- Bersch [13] authors have used SVM classifier to learn which points are the best for grasping. That is to say, the article is also focused on the grasp point detection. Features for each of the points are measured. It is learned which combination of the features was the best for grasping. The situation was simplified by fiducial markers stuck on a shirt surface.
5.5 How the others manipulate cloth

The following actions were used: picking up a shirt from a table (highest point detection) → creating a model (rotating the shirt and creating a point cloud) → estimating current grasp point (uses a position of the markers) → selecting next grasp point (the learned information was used) → computing grasp pose → executing grasp → performing grasp verification → folding (open loop).

- Ramisa [19] tried to find the best grasping point on a shirt as well. The main motivation of this article is to extract as much as possible information about a shirt that lies on the table before a robot picks it up. A “Bag of Features (BoF)” and SVM classifier were used to create a regression model of a general representation of a shirt. The process was as follows: capturing an image of the shirt → dividing it into subregions → computing geodesic-depth histograms (GDH) for each of the subregion → computing BoF on each GDH and stack it into histograms → comparing the histograms with learned regression model to tell which of the regions may contain a collar → grasping the collar (highest point in the proper subregion).

- Hamajima [16] tried to deal with one piece of cloth separation from a pile task and the piece of cloth classification. A color segmentation was used. If the color of the pile was solid, a shadows that appears on the clothes will be used. A model based on connected hems has been used. The decision about which kind of a cloth a robot is manipulating is made by using the information about detected hems and their connections. The model was used for planning cloth unfolding. Hems detection algorithm is given in the article as well.

- The stereo-vision with projected markers was used in Hata [17]. Two densities of markers were used. One sparse markers density is used for quicker orientation in the image. Points of the second higher density are used for creating a cloth 3D model. A towel was used for demonstration. The used process is known from others: highest point detection → picking up → the lowest point detection → re-grasping → ...

- Cusumano-Towner [14] built their algorithm of two basic phases. The first phase is called “Disambiguation Phase”. A probabilistic model (Hidden Markov model [HMM]) was used to determine the identity of a piece of cloth and its configuration. The second phase is called “Reconfiguration Phase”. In the second phase, authors try to convert the model from a known configuration to a desired configuration. The HMM was used in this phase to reduce quickly the uncertainty in the
model state. Next, a cloth simulator based on a strain-limited model was used. Finally their own planning algorithm generated a plan in a reconfiguration phase. The planning algorithm generated a sequence of re-grasp moves to get the cloth article into a desired configuration.

- Miller [22] proposed a new parameterized shape model for clothing. Its main idea is that a piece of cloth can be parameterized by detected landmarks (such as collar, sleeve, ...). A polygon fitted on detected landmarks defines a shape of a piece of cloth. Constraints between the landmarks allow to distinguish between garment classes (shirt, pants, ...) and enable planning of cloth folding. The polygon can be used to perform intelligent manipulation of the clothing article. The constraints in the model are determined by skeletal model of a clothing. The authors create a separate model and a folding sequence for each of the garment classes.

The model has a restriction. It expects that a piece of cloth can be crudely spread out on a flat surface. It assumes that there is a procedure able to spread the cloth.

The article describes the used procedure is as follows. For each known class model → detect expected landmarks (expectation is given by the selected model) → create a polygonal contour encapsulating landmarks → compute a function (called energy function) which describes quantitatively the correspondence between the model and a particular article → optimize parameters such as rotation, translation, scale and deformation to minimize the energy function. After computation of the energy function for each of the classes, select the one with the minimal energy. A selected energy function determines the class model that is the most probably similar to the manipulated object. The folding procedure, which was prepared for each class in advance, follows.

- Wang [24] dealt with sock manipulation and its paring. The main goal was to convert an initial pose of a sock into to predefined configuration. The pairs between socks were found next. The authors started with a garment model that was based on the model of garment shape proposed in Miller [22]. The model was extended to embody local features detected on the cloth such as the cloth micro-texture. The authors described the following key issues:

  - A minimal set of manipulation primitives needed for sock manipulation.
Algorithms used for features computation from garment texture and shape.

The way how to use and train the SVM for feature classification such as location of toes and heel areas.

The global model that considers from three parts: model itself, model cost and a way how to fit parameters.

Two possible ways how to do a sock configuration detection - via an appearance features, via a global model.

A sock pairing algorithm.

The authors performed the task as follows. Measure a feature dataset used for training classifiers → perform the texture classification → recognize a sock configuration → change a sock configuration to avoid sock bunching → repeat previous steps for all socks → pair socks.

6 Conclusions

The main aim of this article was to learn what the others achieved in this area. We also sought the unused space, in which we could bring a novel contribution.

The analyzed articles teach us about several approaches how to detect cloth features such cloth corners, hems and collar detectors. The planning part of most cited work has been rather pragmatic and useful for specific tasks only.

Our main interest was in cloth modeling. Most found modeling approaches were constraint to 3D modeling and rendering. The models were based on physical description of cloth using partial differential equations. Such approaches rely on the finite elements method, which makes them prohibitively computational demanding for our purposes. This is why we seek a new alternative way how to model a piece of cloth.

The model we propose is based on the idea that a piece of cloth or a heap of garments can be represented by a 2D manifold in a 3D scene. Semantically meaningful labels such as sleeve, hem, collar and others can be assigned automatically to pieces of the perceived manifold. Detectors of several features will help in this classification task. A topological map can be created and maintained on top of detected features as well as on top of detected pieces. The topological map will be also used for robot motion planning. The cloth model is suggested as a bottom-up structure of four layers and should be less computationally intensive.
The aim of the diploma project to be finished in about four months is to create a pilot implementation validating meaningfulness of the approach. We leave the planning issues (called reasoning engine earlier in this text) deliberately aside.

References


REFERENCES


