Clothing Simulation of LaundryBot

Wei Wang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-132
http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-132.html

May 30, 2012

Clothing Simulation of LaundryBot

By

Wei Wang

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Engineering

in

Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Prof. Pieter Abbeel, Chair

Prof. Jitendra Malik, Co-reader

Spring 2012

Abstract

Clothes are highly deformable objects. Existing robots are robust in folding a cloth from an easily recognized state, e.g. flat on the table, while they are clumsy in recognizing clothing articles in a previously unseen configuration. We focused on assisting our robot, i.e. the Personal Robot 2 (PR2), on identifying arbitrary configurations of a shirt. We modeled a shirt by printing a number of markers over it and recording the positions of these distinguishable markers. We implemented a computer-vision based system that tracked visible markers on the shirt, and reconstructed the shirt using the world coordinate of each marker. Furthermore, we created a realistic physics-based simulation that accurately models the real world, including gravity, self collision, rigid body motion and so on, to generate instructions to PR2 by taking the configuration of a shirt as input and work out a sequence of actions to take based on existing algorithms. The deliverables of our project indicate that a robot can recognize a clothing article in any arbitrary state with appropriate use of markers and in future the techniques can be applied to related applications.

Acknowledgment

This research project would not have been possible without the support of many people. We wish to express our gratitude to the supervisor, Prof. Pieter Abbeel who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest gratitude is also due to members of the Berkeley Robotics Lab, John Schulman and Ziang Xie without whose knowledge and assistance this study would not have been successful.

Special thanks also to Prof. Jitendra Malik, Dr. Christian Bersch for sharing the literature and invaluable assistance. Not forgetting to our friends who always been there.

Our team wish to express our love and gratitude to beloved families and friends for their understanding and endless love, through the duration of this project.

Table of Contents

List of Figures

List of Charts

1. Introduction

Washer and dryer have come into people's life for a long time in order to automate laundry work while folding still remains as a manual task. Huge efforts have been put onto folding yet no complete and robust system is presented. Robotic manipulation is still comparatively limited compared to human beings as it takes up to 10 minutes for an existing general purpose robot to fold a single T-shirt. The problem occurs because of the lack of a reliable model of clothes and once robot takes an inaccurately shirt configuration, it will follow a wrong path of folding and the whole process will fail.

Clothes are highly deformable, meaning that shape, appearance and other visual properties are very likely to vary due to previous handling and different environments. [1] The biggest problem for a robot is to identify the configuration of a cloth. We aimed at creating a vision based simulation platform to enable the robot to virtually 'see' the shirt and thus determine the configurations and further actions of a folding process. The problem of 'seeing' a shirt poses a challenge to researchers and our solution is to seek help from markers.

Our vision system is based on marker identification using a Kinect and ArToolKitPlus. First, we printed distinguishable markers over the entire shirt, took pictures with Kinect and used ArToolKitPlus to detect and identify every possibly seen marker. Second, we stretched out marker's information, i.e. marker ID and world coordinates, from Kinect and mapped each visible marker to corresponding locations on the shirt. Finally, the shirt will be reconstructed in the simulation platform and this allows the robot to determine its configurations and took further actions.

The simulation platform that we used was the Bullet Physics which has an advantage in modeling soft body objects. We took extremely careful considerations in initializing the simulation in order to accurately model the real world, including gravity, air friction, wind force, soft body self collision, rigid-soft body collision and etc. such that the simulation is realistic and physics based. Since we could model the configurations of a shirt from above vision system, we implemented the whole process of folding a shirt from an existing algorithm by repeatedly grasping the lowest point on the clothes and starting the predefined folding process after finding a known state by the Hidden Markov Model (HMM).

In this report I will focus on the vision system which we developed in the spring semester. Our group had a clear separate of work; I was in charge of the use of Kinect and ArToolKitPlus to detect and identify markers, acquiring world coordinates of visible markers and give them to Ruben. Ruben then reconstructed the shirt in Bullet Physics because he is more familiar with the simulation platform. Ruben also worked on some supplementary functions of marker identification, i.e. unwarping the warped markers (section 3.4.2) and the moving camera scheme (section 3.4.1) in order to overcome the limited resolution of Kinect. Siyu looked into the possibilities of using invisible markers (section 3.4.3) so that our deliverables have more market applications. In the fall semester our group has successfully demonstrated the use of Bullet Physics to simulate the folding process given the configuration of a shirt.

The remainder of this report is organized as follows. In section 2 I will give a literature review on related researches to my part. In section 3, I will introduce and explain our vision system in details. In section 4 I will show our experimental results as well as discussing any relationship and improvement over literature. In section 5 I will present our conclusion and discuss future work to achieve a better result.

2. Literature Review

2.1. Bringing a shirt to a known configuration

The process of folding a shirt from a known state has been well developed. Researchers have put much effort on bringing a shirt from any arbitrary state to a known state in order to apply the folding algorithm. The state-of-the-art technology in this field comes from Marco Cusumano-Towner et al. [1] whose solution consisted of a disambiguation phase followed by a reconfiguration phase. In the disambiguation phase the robot repeatedly grasped the lowest point of the cloth, took observation and chose a most likely configuration using HMM. In the reconfiguration phase, taking the output configuration from the previous phase as input, the robot planned a sequence of actions to take and brought the cloth into the final desired state. The flaw in Marco's model was the speed. The disambiguation phase could take up to 10 minutes since it kept looking for a known configuration by cues from 2D perception. The primary goal of our project is to solve the speed problem.

We also referred to work by Jeremy Maitin-Shepard et al [2] as they successfully developed a system that allowed a generous purpose two-armed robot to fold a towel. Their approach was to start from a random configuration, go through a sequence of vision-based actions and find the edge and thus the four corners of a towel based on distinct features of each part. After the robot detected the corners, it became easy to fold the towel and bring it to a target location. This system has a similar flaw to Marko's since the robot needs to go through a set of predefined actions to configure the towel and this is potentially time consuming. Furthermore, Jeremy stated that this system worked for towels only because features of a shirt were more complicated with sleeves and necks. We improved the situation by putting markers on a shirt and thus turning the shirt into an array of markers.

2.2. Configuring a shirt using markers

The marker technology involved in our project is similar to the work presented by Christian Bersch et al. [3] Christian printed fiducial markers containing digital information as IDs as described in [4] over the entire shirt. ArToolKitPlus [5] was adopted by Christian as a tool to detect and identify the markers. In Christian's system, the robot grasped a point of the shirt and lifted it so that the shirt was in a corn shape. The robot then rotated the shirt slowly, searched for visible markers, mapped the locations of each marker to some known configurations and changed grasp points until the shirt reached a known configuration. Our system is very similar to Christian's except that ours did not involve any grasping of the shirt because there was no simulation in Christian's system so they had to manipulate the shirt until the permutation of markers were matched to a known configuration. Our simulation was reliable so we could transfer any state of the shirt into the simulation and determine a sequence of actions accordingly. The advantage of avoiding manipulations of a shirt was a significant increase in speed since computer simulation is much faster than robot manipulations.

3. Methodology

In order to accomplish our final goal, we firstly need to print distinguishable markers over the entire shirt. We chose BCH markers because it offered over 1000 distinct patterns which were more than enough for our project.

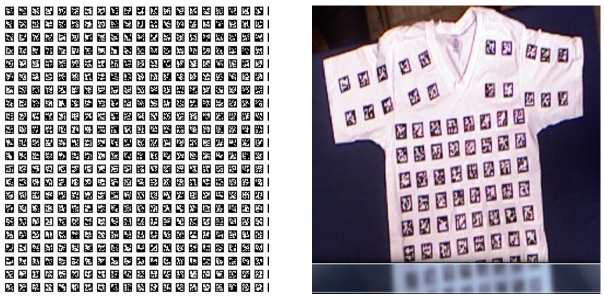

Figure 1: BCH markers (left) and the marked-up T-shirt (right).

Notice that due to time constrain and technical issues, we did not print markers onto the shirt. Instead, we printed them on the paper and pasted them onto the shirt.

The entire process of our project is simplified and shown in the following flow chart.

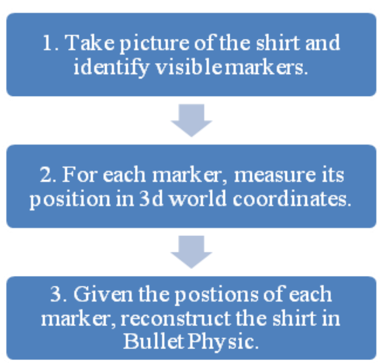

Chart 1: The entire process of our project.

The Bullet Physics was the simulation platform that we adopted in fall semester. My teammate, Ruben, was in charge of step3, the reconstruction step in Bullet while I worked on the identification and position measurements of each marker. In order to achieve this goal, the primary hardware involved was Kinect for Xbox 360 and software I used were ArToolKitPlus, OpenCV and official Kinect SDK from Microsoft.

2.3. ArToolKitPlus

In order to detect and identify markers, as mentioned in previous sections, we used the ArToolKitPlus. This tool was able to detect BCH markers from a grayscale or binary image by pattern matching schemes. Since a number of researchers have tested and proved the stability of this tool, we simply took a picture and the software returned a list of detected markers together with their IDs, pixel locations and pose matrices in reference to the camera. I made use of the IDs and pixel locations combined the depth information from the Kinect to determine the world coordinates of each marker. The pose matrices were discarded because they the accuracy was not enough, according to Christian Bersch.

Definition:

To avoid confusion, I define ‘detecting’ a marker and ‘identifying” a marker here.

By ‘detecting’ we mean the ArToolKitPlus knows there is a marker but is not sure which marker it is. By ‘identifying’ we mean ArToolKitPlus asserts the ID of a detected marker. The difference is because ArToolKitPlus can determine the existence of a marker by roughly matching a blurred or unclear area of black and white grids to the saved patterns while in order to tell exactly which marker it is, it has to see every single piece of grids and apply pattern matching schemes. The gap between ‘detection’ and ‘identification’ comes from the distance and resolution of the camera.

2.4. Adaptive thresholding

ArToolKitPlus is sensitive to lighting changes because it uses a binary image as input and identifies the unique permutation of black/white grids within each marker. The brightness difference between a black pixel and white one tended to vary drastically when lighting changes and thus we applied adaptive thresholds to the image.

The adaptive thresholding technique assigned a binary value to each pixel depending on local brightness around each pixel and thus significantly reduced the negative effect of lighting.

2.5. Kinect

Kinect is a device by Microsoft which is designed for motion sensing as an add-on peripheral for the Xbox 360 console. [6] Now many researchers use Kinect as a stereo camera to grab color images as well as the depth maps. The reason to use a Kinect in our project was because it could return 3D coordinates of markers in color images. Basically we passed the adaptive thresholded image into ArToolKitPlus and it identified markers as well as returning the pixel locations of four corners of each marker. Then we calculate the 3D coordinates of a marker by first finding the distance of the marker with respect to Kinect by averaging pixel depths inside the four corners of a marker followed by a perspective projection.

The challenge was that Kinect had a minimum working distance of around 90cm within which distance measurement would not work. On the other hand, since Kinect had a limited resolution, i.e. 640*480, the shirt could not be placed too far from Kinect. In fact, if the distance was beyond 80cm, the markers could not be identified by ArToolKitPlus. So the range of distance from the shirt to Kinect was rather limited and we could not identify the marker and measure 3D coordinates at the same time. Thus we designed the following 'moving camera' scheme to handle this problem. Since it was Ruben who mainly worked on this scheme, I put it under 'related work by our group' section.

2.6. Related work by our group

2.6.1. The moving camera scheme

Referring to the following flow chart, we basically started from a small distance to the shirt to let ArToolKitPlus identify every marker. Then we slowly moved the Kinect away from the shirt. At the same time we kept tracking of every marker that ArToolKitPlus detect, until it reached a point where Kinect was able to return depth information.

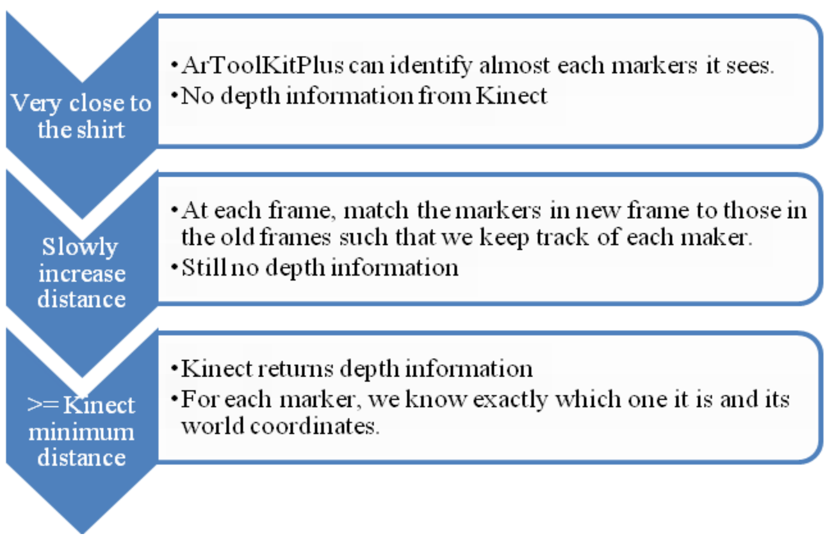

Chart 2: The flow chart of the moving camera scheme.

With this scheme, our system could actually track every marker and finally get a list of visible marker together with their world coordinates. This list would then be inputted into the Bullet Physics to reproduce the shirt. There was no speed problem here since this scheme only took around 20 seconds which was much shorter any robotic manipulation.

2.6.2. Unwarping a warped marker

A marker is easily warped when the shirt is not flat on the table and thus leads to incorrect identifications. My teammate, Ruben, developed an algorithm to unwarp a warped marker based on an Affine transformation between the locations of four corners of a marker and the locations of a standard square. Each pixel of the warped marker went through the transformation and formed a new square marker. Using this technique our system would be able to identify as many markers as possible.

2.6.3. Invisible markers

As for the requirement of Master of Engineering program, we looked into the marketing aspect of our marker technology. Apparently few customers are willing to buy a T-shirt with BCH markers on it. My teammate, Siyu, discovered the use of some interesting invisible marker that could potentially put our marker technology into the market.

3. Results and Discussion

4.1. ArToolKitPlus Result

As mentioned in the previous section, ArToolKitPlus was chosen to detect and identify markers. The following three graphs showed the image taken by Kinect from different distances.

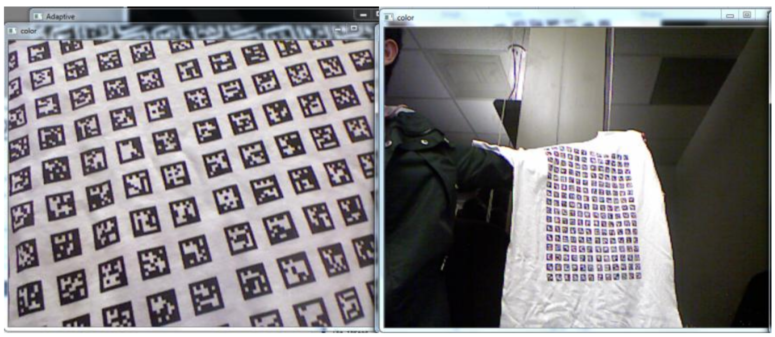

Figure 2: Photos from Kinect at a close distance to the T-shirt (left) and a far distance (right).

Results:

	Left	Right
Detect	62	120
Identify	57	3

As indicated by the results, when the clothes were close to the camera, ArToolKitPlus could detect and identify every marker visible markers. As the distance became larger, it could still detect many markers but hardly identify any. That was the reason we implemented the moving camera scheme.

4.2. Final Results

We hereby present the experimental results for the overall system, including marker identification by ArToolKitPlus, the moving camera scheme, upwarping and reconstruction in Bullet Physics. In this report we will show figures of four cases and discuss them one by one.

Case 1: The T-shirt was flat on the table.

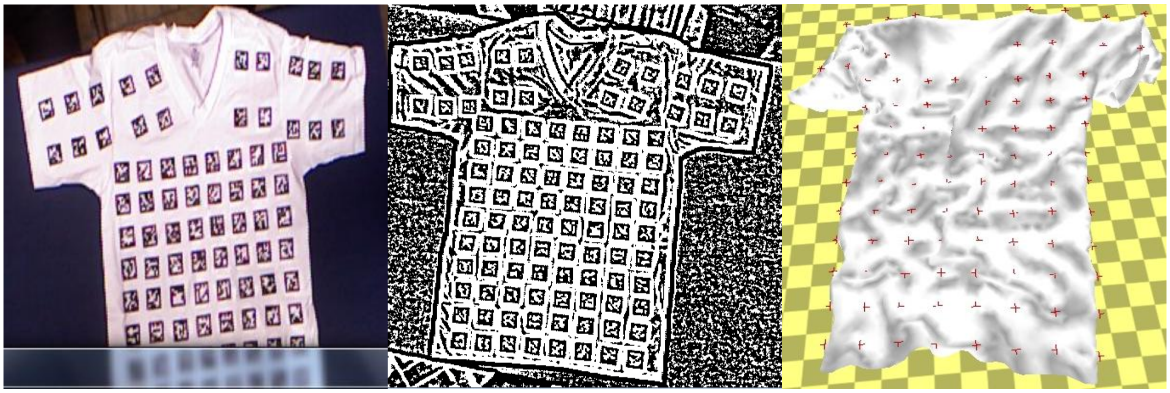

Figure 3 (From left to right): the T-shirt in the real world; the T- shirt in Kinect's view with adaptive thresholds; the reconstructed T-shirt in simulation.

As we can see from the rightmost figure, the shirt was perfectly reconstructed with a clear distinction between body part and sleeves. The array of crosses demonstrates the pattern of markers with each cross represents a marker.

Case 2: The T-shirt had a fold which is pointed by an arrow in the leftmost figure.

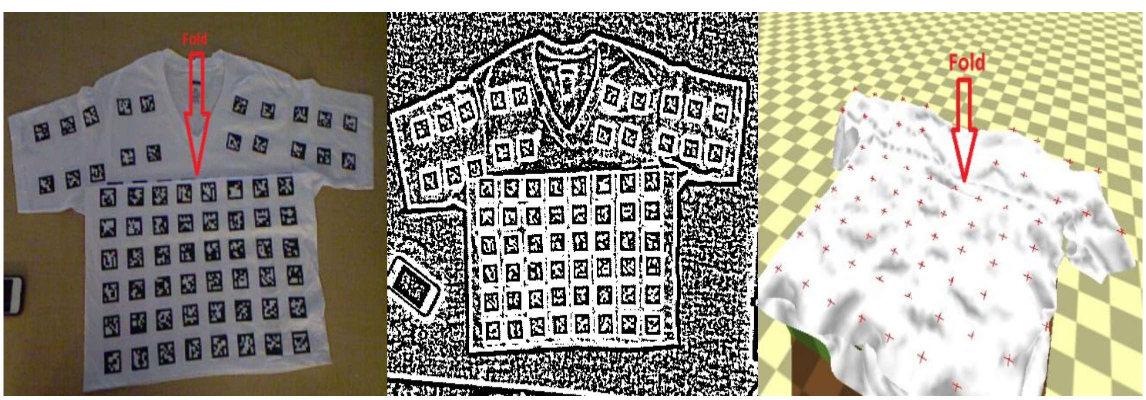

Figure 4 (From left to right): the T-shirt in the real world; the T- shirt in Kinect's view with adaptive thresholds; the reconstructed T-shirt in simulation.

In rightmost figure, an arrow is pointed to the location of the fold. As indicated by this result, our system is able to accurately model the T-shirt with the exact location of the fold.

Case 3: T-shirt had a fold which is pointed by an arrow in the leftmost figure.

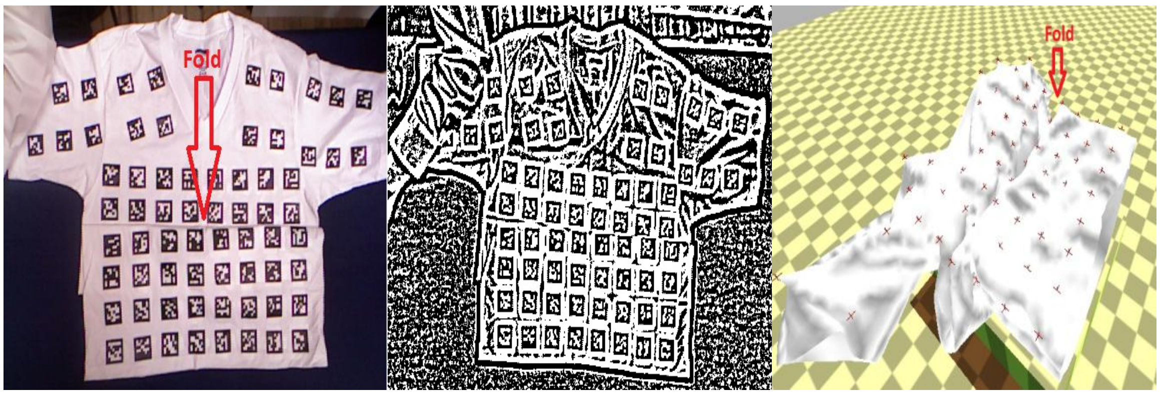

Figure 5 (From left to right): the T-shirt in the real world; the T- shirt in Kinect's view with adaptive thresholds; the reconstructed T-shirt in simulation.

In this case the reconstructed T-shirt was viewed from another angle. In rightmost figure an arrow is pointing to a gap between two parts of the T-shirt. The reason for the existence of this gap was that the Kinect could not see the markers under the fold and our system simply ignored these markers. The simulation engine allowed a piece of cloth to drop by gravity if no markers were found over that piece. So some parts of the T-shirt were fixed in the air while other regions that had no markers detected dropped creating a gap in the image.

The result indicates that our system is able to reconstruct any markers it sees while it cannot predict the locations of undetected markers. This is reasonable since human beings cannot tell the configuration of the piece of cloth under the fold, either.

Case 4: The shirt was in an arbitrary configuration which was more likely to be seen in a real folding process.

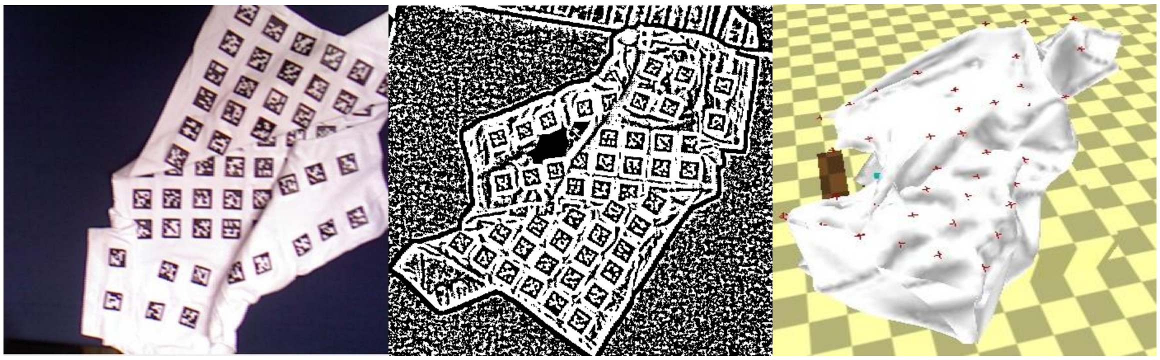

Figure 6 (From left to right): the T-shirt in the real world; the T- shirt in Kinect's view with adaptive thresholds; the reconstructed T-shirt in simulation.

As shown in the rightmost figure, the reconstruction is quite successful as we can easily tell the location of each unfolded part of the clothes.

Overall, performance of our system is stable and robust with any arbitrary configuration of the T-shirt. Our primary goal is to increase speed of recognizing a configuration. In our experiments, each case only takes up to 1 minutes and most of the time was spent on reconstruction since the simulation needed to handle air-friction, gravity, and cloth self-collision at the same time. This portion of time can be reduced with faster processors. The time can be further reduced with the moving camera scheme eliminated if a binocular stereo camera with higher resolution is used.

4.3. Invisible markers

The left figure shows the marker under a UV light and the right figure shows what it looked like in the view of Kinect after applying a binary threshold to the image.

Figure 7: the invisible markers under UV light in the real world (left) and in the view of Kinect (right).

These markers were initially invisible to human beings. Under the exposure to UV light, the markers were able to be identified by ArToolKitPlus after applying some simple image processing technique. The result indicates that it is possible to replace BCH markers by invisible ones and make our technology more market friendly. Due to time limitation, we did not print invisible markers over the entire shirt and conducted further experiments.

4. Conclusion

The project gives us an opportunity to gain insights of cutting edge robotic manipulation technology. Through the research, I acquire plenty of computer vision knowledge including stereo processing and pattern recognition of visible markers. I also gain hands-on experience in working with a stereo camera such as the Kinect. To go further in this project, the solution to unidentified markers versus minimum working distance of a stereo camera should be carefully studied.

Our team has successfully implemented a system to identify markers in the real world and use the 3D coordinates to reconstruct a shirt with printed markers in the Bullet Simulation Engine. Since my teammate has developed an efficient algorithm to reconstruct the shirt in the simulation platform with found markers, the weakest point in our project is the absence of a practical approach to identify every marker on a shirt. This drawback may lead to unsuccessful recognition of the states of a clothing article because of missing markers.

The technology involved in this project has a huge number of potential applications. The simulation engine can be used to let researchers do experiment and test their algorithms. The interface we developed between the real world and simulation can be applied to any vision-related commercial goods, such as any camera, target capturing system, the cleaning robot – Roomba and so on. Our discovery in invisible markers could potentially be a practical solution in robot cloth manipulation.

To future researches in this field, I recommend the following:

1) Try to use a stereo camera with higher resolution than Kinect to avoid moving camera scheme. The moving camera scheme is designed and implemented because of the limitation of our hardware. A better stereo camera can save both time and space as it will require taking only one photo and do not have to track markers frame by frame. Kinect has a separate stereo system and RGB camera. Using a binocular camera is enough for this application and can also avoid errors in mapping pixels from color image to depth image.
2) Try to print invisible markers which can only be detected under a UV light instead of BCH markers over the T-shirt. If the use of invisible markers is proved to be technically functional and can fit into our system, the market size of the marker technology is going to increase drastically.

Reference

[1] A. S. S. M. J. F. O. P. A. Marco Cusumano-Towner, "Bringing Clothing into Desired Configurations with Limited Perception," in *IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (ICRA)*, 2011.

[2] M. C.-T. J. L. a. P. A. Jeremy Maitin-Shepard, “Cloth Grasp Point Detection based on Multiple-View Geometric Cues,” *Robotics and Automation (ICRA)*, 2010.

[3] B. P. S. K. Christian Bersch, “Robotic Cloth Manipulation using Learned Grasp Evaluations,” *RSS Workshop on Mobile Manipulation: Learning to Manipulate*, 2011.

[4] M. Fiala, “ARTag, a Fiducial Marker System Using Digital Techniques,” *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),* vol 2, ISBN 0-7695-2372-2, pp. pp.590-596, 2005.

[5] D. W. a. D. Schmalstieg, “ARToolKitPlus for Pose Tracking on Mobile Devices,” *Computer Vision Winter Workshop*, 2007.

[6] “Kinect from Wikipedia”. Available: http://en.wikipedia.org/wiki/Kinect.