DEXTEROUS MANIPULATION PLANNING FROM HUMAN DEMONSTRATION

Phongtharin Vinayavekhin

Thesis Supervisor: Katsushi Ikeuchi

Submitted to
the Graduate School of the University of Tokyo

in Partial Fulfillment of the Requirements
for the Degree of
Master of Information Science and Technology
in Information and Communication Engineering

February 2009

Abstract

In recent years, many multi-finger robotic hands have been designed and developed in order to emulate a human hand. The human hand has many sophisticated a lot of complicated dexterous skills that human themselves sometimes did not notice, for example an ability to grasp and manipulate tiny objects, a capability to regrasp an object without releasing and so on. To mimic or imitate these kind of dexterous skills to high degree-of-freedom mechanic hands, a planning is required and it is often referred as Dexterous Manipulation Planning (DMP).

In this thesis, DMP is explored in context of learning object manipulation using programming by demonstration (PbD) framework. The idea of this paradigm to plan a dexterous manipulation by letting robot observes a stream of human demonstration. An advantage of PbD is that the planning discovered will also include human purpose in manipulation. This is corresponding to previous research which suggested that human has particular reasons on how they grasp or manipulate objects.

This thesis proposes a novel method to segment a stream of human manipulative movement. This is a preliminary step to plan and mimic dexterous manipulation ability to robot hands in a PbD framework. Human manipulative movement is observed by a data glove, which provides information of 18 joint angles of human hand. Then a highly dimensional joint space is reduced to a lower dimensional space using principle component analysis. A segmentation is done based on two assumptions: First, a contact relation between hand and object would change when a coordinative movement of all joint angles of the hand changes it direction. Second, a coordinative movement of all joint angles are an approximately linear line in a reduced dimensional joint space.

To test and apply the proposed method three manipulative movements of human is considered. Two of which are the repetition of same simultaneous movement. The other is more complex manipulation of a pen-like object called interdigital step. The results are compared and discussed with another previous research.

Acknowledgments

First of all, I would like to express my sincere gratitude to my thesis supervisor, Prof. Katsushi Ikeuchi, for his consideration and supervision. His advices and comments during each meetings always have a great impact on my research. I also would like to thank him for a great research environment provided in his laboratory.

My appreciation goes to Dr. Shunsuke Kudoh for his direction. His suggestion always helps me decide what is important and should be considered as a priority. I also admire his strategy of problem solving, which I have absorbed while working under him. I believe that these gifts and skills would be invaluable to my future.

Many thanks go to Dot-Chan who had always broken down, but did recovered by the time I need it. Technical skills acquired during his maintenance and reparation (which is not quite completed yet) are beyond price. I also would like to thank Dr. Koichi Ogawara and Mr. Yoshihiro Sato for providing me a robot platform and their great assistances in repairing Dot-Chan.

It is my great pleasure to be part of CVL laboratory. I would like to thank Okamoto-san and Mawo-san for helping me survive in Japanese culture, Miti-san and Manoj-san for friendship and encouragement, Bjoern for your brilliant debian skills, Ninta for last minute reading, all secretaries for managing my official documents, and all the other CVL members for being great hosts.

Life in Japan would be totally boring without Soshigaya and MEXT 2006 community. All seniors, volunteers, and friends should have all these credits.

I also would like to acknowledge Japanese Monbukagakusho for support that made it possible for me to study in Japan.

I am also very grateful for my parents and sister, whom I dedicate this thesis. Without your love, support and encouragement I would not have come this far.

You'll never walk alone.

CONTENTS

1 Introduction **1**
- 1.1 Programming by Demonstration (PbD) . . . 2
- 1.2 Object Manipulation using PbD . . . 3
- 1.3 What is This Thesis About? . . . 5
  - 1.3.1 System's Breakdown . . . 6
  - 1.3.2 Segmentation's Criteria . . . 10
- 1.4 Thesis Organization . . . 12

2 Related Works **13**
- 2.1 Characteristic of Prehensile Movements . . . 13
- 2.2 Intermediate Grasp State Extraction . . . 18
  - 2.2.1 Hand Posture Based Approach . . . 18
  - 2.2.2 Tactile Sensor Based Approach . . . 19

3 Intermediate Grasp States Extraction **21**
- 3.1 System Formation . . . 21
  - 3.1.1 Defining Grasp Taxonomy . . . 22
  - 3.1.2 Data Acquisition System . . . 24
  - 3.1.3 Task Primitives Segmentation & Recognition . . . 24
- 3.2 The Proposed Method . . . 25
  - 3.2.1 Dimensional Reduction using PCA . . . 27
  - 3.2.2 Extract Grasp States using Curvature . . . 29

4 Experiment and Evaluation **33**
- 4.1 Experimental Data and Setups . . . 33
- 4.2 Experimental Results . . . 36
  - 4.2.1 Experiment 1 : Dynamic Tripod of the Pen . . . 37

4.2.2 Experiment 2 : Rock the Rubik Cube 44
4.2.3 Experiment 3 : Interdigital Step of the Pen 51
4.3 Discussion 59

5 Conclusion **61**
5.1 Future Works 62

Bibliography **65**

LIST OF FIGURES

1.1 Pre-defined task primitives in particular task 2
1.2 A Task Model 3
1.3 Example of possible task model. 4
1.4 Skill parameters extraction of in-hand dexterous manipulation 5
1.5 Object manipulation system's breakdown 7
1.6 One possible example of grasp transitions graph for Object A 9
1.7 Transition of contact relation in assembly task; Takamatsu [1] 10
1.8 Criteria for segmenting a stream of hand postures demonstrating a dexterous manipulation 11

2.1 Power grasp (l) and Precision grasp (r); Napier[2] 14
2.2 Grasp taxonomy used by Cutkosky's expert system[3]. 16
2.3 Virtual finger and three basic oppositions (figure [4]) 17
2.4 Hand postured in reduced joint angle space (2D) 17
2.5 Temporal segmentation for object manipulation; Kang [5] . . . 18
2.6 Experimental results of Zacksenhouse's segmentation technique 19
2.7 Data acquisition and experimental result; Kudoh[6] 20

3.1 Object manipulation during painting activity. 22
3.2 Grasp used for drawing and painting 23
3.3 Grasps used for measuring 23
3.4 Grasp used for other purposes 24
3.5 System setup of a data acquisition system 25
3.6 Experimental results of static grasp recognition using SVM . . 26
3.7 Example of trajectory of hand postures in 3D-eigengrasp space 29
3.8 Simplified version of how to calculate curvature 31
3.9 Curvature of trajectory in figure 3.7, in 10 different scales . . . 31

4.1 Sketch of dynamic tripod and rock movements; Elliott[7] . . . 33
4.2 Sketch of interdigital step movement; Elliott[7] 34
4.3 Exp. 1: Snapshot of hand movements in frame 69^{th}-150^{th} . . 37
4.4 Exp. 1: Accumulative variance accounted by each eigengrasps 38
4.5 Exp. 1: Projected trajectory and its curvature 39
4.6 Exp. 1: Example of some extracted intermediate grasp states . 40
4.7 Exp. 1: Projected trajectory in other experimental setups . . 41
4.8 Exp. 1: Comparison of curvatures with all other experimental setups 43
4.9 Exp. 2: Snapshot of hand movements in frame 63^{rd}-144^{th} . . 44
4.10 Exp. 2: Accumulative variance accounted by each eigengrasps 45
4.11 Exp. 2: Projected trajectory and its curvature 46
4.12 Exp. 2: Example of some extracted intermediate grasp states . 47
4.13 Exp. 2: Projected trajectory in other experimental setups . . 48
4.14 Exp. 2: Comparison of curvatures with all other experimental setups 50
4.15 Exp. 3: Snapshot of hand movements in frame 16^{th}-124^{th} . . 51
4.16 Exp. 3: Accumulative variance accounted by each eigengrasps 52
4.17 Exp. 3: Data loss due to trajectory smoothing 53
4.18 Exp. 3: Projected trajectory and its curvature 54
4.19 Exp. 3: All extracted intermediate grasp states 55
4.20 Exp. 3: Projected trajectory in experimental setup 4, when considered two different phrase-planes 56
4.21 Exp. 3: Comparison of curvatures with other experimental setups 58

5.1 A difference in length after projection 63

LIST OF TABLES

2.1 Kamakura's Grasp Taxonomy [8] 15

CHAPTER 1 Introduction

There is a prospect that in the near future ROBOT will become part of our environment, and play important role in human society. Robots which once only existed in the novels, now have gained more of their existence in human everyday life, e.g. a robotic vacuum cleaner, or even a little toy. Human has expected more from robots. Human hopes that someday robots will be able to work alongside and assist them in their environment.

There are so many challenges for enabling the robots to work in human environment [9]. One of the important challenges is how the robots would be dealing with a human environment. In real situation, human circumstance can be very complicated for the robots, when compared to the factories or laboratories where everything can be controlled. A traditional approach where knowledges are pre-programmed confines robots to limited abilities and intelligences. This restriction would result in an incompetence of robots to fully participate in dynamic environment without further intervention, once they are manufactured and leave the factory. Therefore, an alternative method to teach or program a robot is required.

One of the approach is to let the robots learn from the natural statistics of human environment. Programming by demonstration (PbD) is an approach to teach robots to learn particular tasks by showing it an example. It has been proved to effectively teach a robot in many kind of applications, range from an assembly task in factory [10] till teaching a humanoid robot to dance [11]. In this thesis, we are interested in teaching a robot an object manipulation task, which can be seen very often in human daily life.

1.1 Programming by Demonstration (PbD)

To teach a particular task to a robot using programming by demonstration framework, we must first define all TASK PRIMITIVES for the task. Task primitive tells the robot what to do. In other words, task primitives are the actions which can be understood by the system.

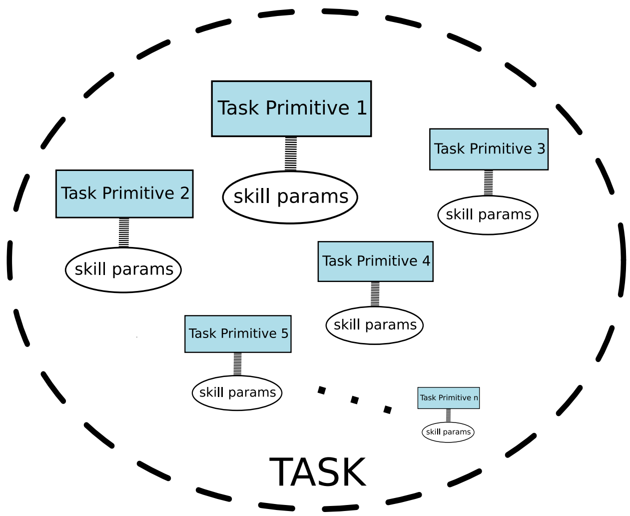


Figure 1.1: Pre-defined task primitives in particular task

TASK is defined as a sequence of task primitives. Teaching a task to a robot by human demonstration can be referred as a process of recognizing and extracting the sequence of task primitives out of the stream of human demonstration.

For each task primitives, there are always corresponding SKILL PARAMETERS. Skill parameters tell the robot how to do or perform a corresponding task primitives. Generally speaking, in programming by demonstration skill parameters should be able to recognize and extract out of the stream of human demonstration as well. However, there are also some skill parameters of some task primitives that might been discovered from other methods.

Once we had recognized and extracted a sequence of task primitives and the corresponding skill parameters from human demonstration, the result has to be stored or written in some kinds of representation, which can be understood by the system. This representation is called ABSTRACT TASK MODEL, or most of the time only referred as TASK MODEL.

In other words, *task model* is a representation of *task* in programming by demonstration system. It is composed of two important elements: a sequence of *task primitives*, and a corresponding *skill parameters*. All of which are recognized and extracted out of a stream of human demonstration.

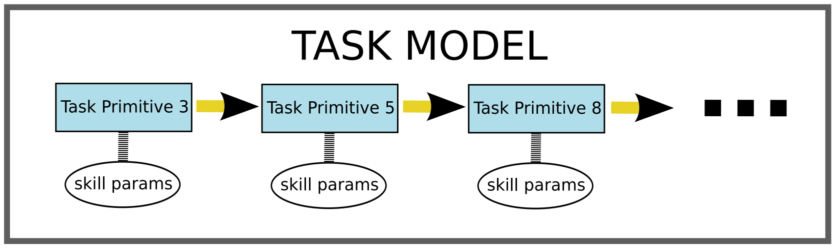


Figure 1.2: A Task Model

1.2 Object Manipulation using PbD

Object manipulation is one of TASK appears everywhere in human everyday life, e.g. in the kitchen where human cooks, in the office where human works, or even on the dinning table where human eats. Teaching object manipulation by programming by demonstration has recently been a active research in robotic filed. Domain of object manipulation task normally limits to activities which can been seen in everyday life. Although it is not definite which task primitives should be defined in everyday object manipulation, one of the approach is to defined task primitives as *grasping*, *manipulating*, and *releasing* [5].

In this context, grasping refers to a hand movement starting from a empty hand move toward an object, until a hand have a good grasp of an object. Releasing refers to a hand movement starting from a hand holding an object with a type of grasp, until a hand releases the object and has nothing within it.

Manipulating task primitive is quite complex and sometimes has a very broad meaning. Manipulating may referred as any hand (or arm) movement which cause an object to move, likes pushing etc. However, in this context we would like to refer to the manipulating as a prehensile movement of a hand that cause an object to move, especially those with a dexterous capabilities.

By defining our task primitives like those defined above, we can make sure that, if a grasped object is about to change to different object, there will always be a releasing task primitives in the middle. In other words, in our system a hand has to release the object it currently grasped, before grasping another object.

There are many manipulative prehensile movements of a human hand. Although some approaches have tried to classify them [7][12], the definition of manipulative prehensile movements are still not definite, and correspond to our interest. Therefore, in our study, we further define manipulating task primitives into more details based on their basic skill parameters: its initial grasp, its final grasp, and its grasped object. Based on this, our manipulating task primitives can be classified into two different groups.

1. manipulating which has same type of initial grasp and final grasp. In other world, type of grasp does not change during manipulation. Its main intention is solely to move a grasped object together with the hand as rigid as possible. We will refer to this type of task primitives as *homogeneous manipulation* [5] task primitives.

2. manipulating which has initial type of grasp different from final grasp. Its main intention is to change a type of grasp currently employed at the object, not to move a grasped object. However, to accomplish such movement, a object transporting may also occur. We will refer to this type of task primitives as *dexterous manipulation* or *regrasp* task primitives.

By mention this, we mean that manipulating task primitives which employed *more than two* type of grasps, will be divided until each of them only contain one or two type of grasps the whole time.

Besides, as we mention *type of grasps* earlier, we would like to define it more concrete here. In this context, type of grasps defines based on an object it grasps. For a particular objects, a *finite* number of static hand postures will be assigned as all type of grasps. The decision on this might be decided by system designer himself/herself, or extracted from a stream of human demonstration. This is done based on the assumption that for particular category of similar objects, human has specific type of grasps for it [8].

Furthermore, on the system as a whole, when many objects are considered, two type of grasps from different category of objects *may* be defined as same grasp, even though their posture may be a little different because of an object it hold. We leave this as a choice for system designer, who would have to build a dynamic grasp recognition system.

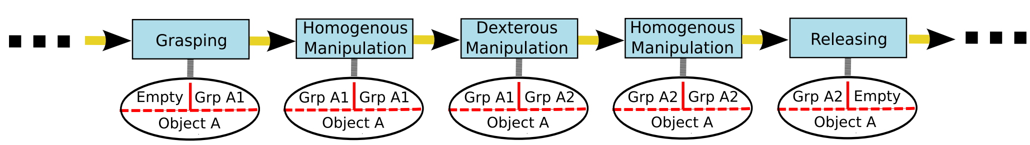


Figure 1.3: Example of possible task model.

One possible task model is shown in figure 1.3. An example task model starts with a hand grasping *Object A* with grasp *Grp A1*. Then the hand manipulates *Object A* while using grasp *Grp A1*. After a while, the hand changes a from grasp *Grp A1* to grasp *Grp A2*, while holding *Object A*. Grasp *Grp A2* is used to move *Object A*, before finally decide to releases *Object A*, and so on . . .

1.3 What is This Thesis About?

Kang [13][5] has developed proposed a approach to program object manipulation by human demonstration. He has developed a system which extracts all skill parameters of the task primitives, which are necessary for mapping to the target robot hand. However, in his task model, only homogeneous manipulations were considered.

In our study, we move one step further. We take into account a dexterous manipulation, a more precise term would be an *in-hand dexterous manipulation*. Considering basic skill primitives of dexterous manipulation task primitives defined in figure 1.3: a grasped object, initial and final grasp type, these parameters are telling a robot what to do, but not really how to do. In other words, they contain insufficient information for target robot hand to automatically execute the dexterous manipulation.

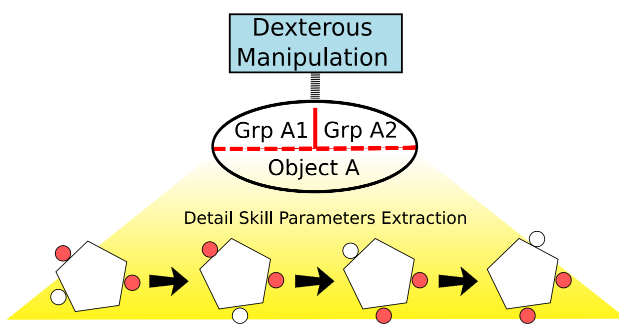


Figure 1.4: Skill parameters extraction of in-hand dexterous manipulation

While there are so many research proposed a method to solve dexterous manipulation problem [14], we decide to stick and follow our main approach, programming by demonstration. This means that we prefer to recognize and extract skill parameters of dexterous manipulation task primitives by observing a stream of human demonstration, or in more generic term, dexterous manipulation planning (DMP) from human demonstration.

Why PbD? . . . not Analytic Approach

This is a frequency asked question. Which approach is more appropriate to teach object manipulation for multi-finger robot hand? Is not there are a lot of researches about dexterous manipulation in an analytic approach? We also do not know the answer exactly on these questions. However, our decision on using programming by demonstration is done based previous literatures in grasp choice, and dexterous manipulation planning.

A study on static prehensile movement of human hand by Napier [2], suggested that one of important factors that influences human which type of grasp to use, is an intended activity. He raised an example of human grasping a wooden rods, and suggest that between two possible type of grasp, which will be adopted by human depends solely the purpose of use of that wooden rods.

Cutkosky [3] has similar statement on this. He built an expert system, which decided semi-automatically on which grasp should be utilized, particularly in manufacturing task. The system worked well based on machinists feedback, which gave an information about task requirement and object shape to the system. Cutkosky also gave some concerns about a performance of analytic approaches, which based on many estimated physical model, outside of the controlled-environment laboratory.

When someone wants to grasp a knife from the table, it is obvious to grasp at the dull side, not at the sharp side. If we use an analytic approach, this would definitely be one of the constrain we have to consider. With the same manner, when someone wants to regrasp a knife, they have to choose a path of the hand which make sure that contact points are not at the shape part, or a path which results in an object trajectory that would not hurt anyone. These conditions also have to be considered as constrains on analytic path planning as well. These are just an simple example, but imagine when we consider everyday object manipulation task, how many these constraints would increase? This assumption is actually correspond to one literature by Kudoh [6], which claimed that in manipulation in everyday life, it is important to imitate the type of grasp as well as to imitate the motion of a manipulated object.

1.3.1 System's Breakdown

Our system model may look very similar to Kang's system [13][5], but taking into consideration a dexterous manipulation has changed some characteristics of the system. As a result, some components already proposed in the previous system may not be able to use here. In this section, we list all necessary components we need, and give more informations to the components, which need to be revised.

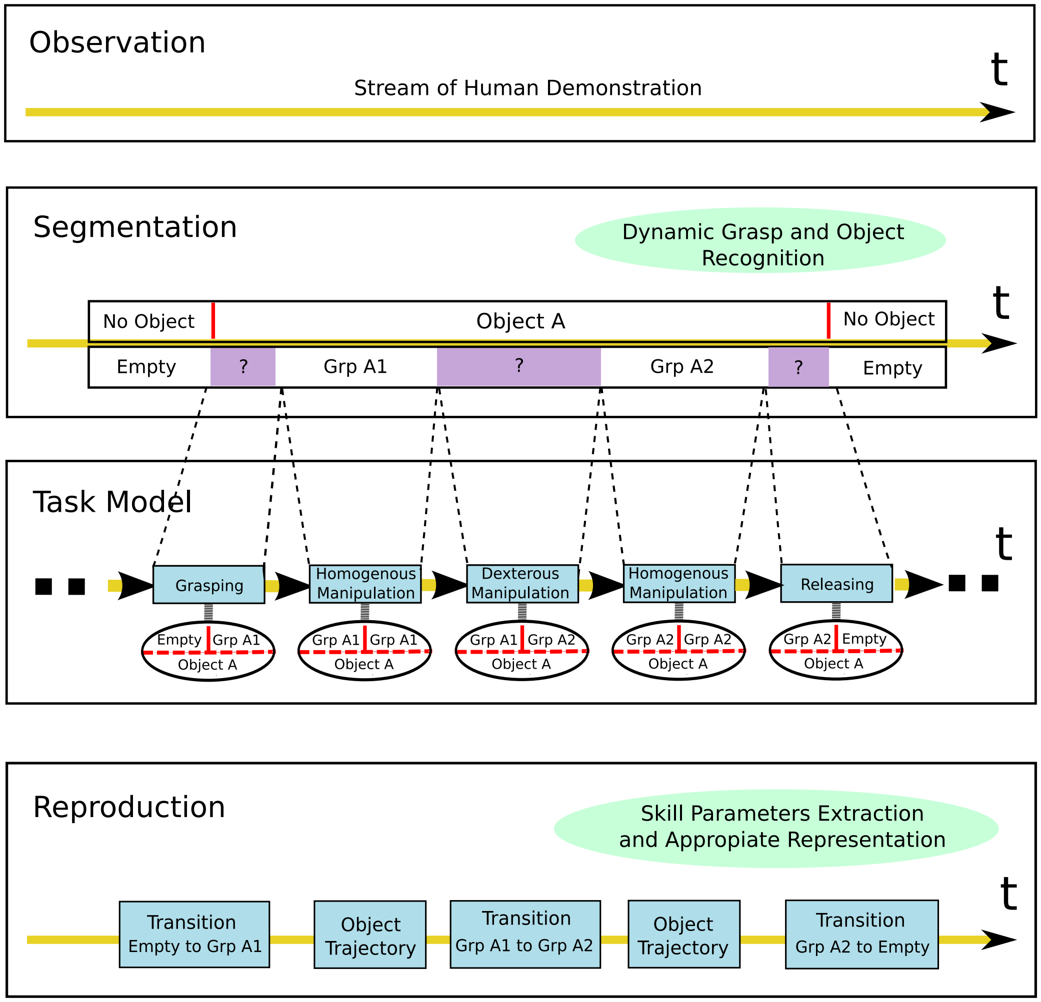


Figure 1.5: Object manipulation system's breakdown

Figure 1.5 shows all necessary components in our system. It starts from observing human demonstration using data acquisition components. This can range from a motion capture system of a hand, to a real-time stereo camera system that can reconstruct a hand model. At a very least, the observed data should consist of a hand posture, a hand location, and an object location. If an object considered is a rigid body object, its geometry may be observed prior to the demonstration.

Once a data is captured, the system must be able to extract sequence of predefined task primitives out of the stream of data. The present of task primitives can be referred from their basic skill-parameters, which are their initial grasp, final grasp, and the grasped object. Therefore, to recover a sequence of task primitives, or in other words a task model , we need to segment a stream of captured data and recognize type of grasps together with an object, or, in short, dynamic grasp and object recognition system.

Bernardin [15] proposed a method to recognize continuous human-grasp sequences from human demonstration. With a use of hand shape and contact point information which acquired from data glove and tactile sensors, he managed to segment and classify human grasp into pre-defined grasps using hidden Markov model recognizer. Although in our system, pre-defined grasps are defined based on each particular objects or each particular category of objects, once the grasping object is recognized, the method can be applied easily.

After grasps and object are recognized, we can build task model based on that information. To reproduce whole object manipulation in target robot system, further skill parameters for each task primitives is necessary. For instance, a grasp planning is required to reproduce grasping and releasing task primitives, object trajectories is needed if a homogeneous manipulation is about to implement. In our study, we proposed a method of dexterous manipulation planning from programming by demonstration, which are required skill parameters when reproducing dexterous manipulation task primitives.

In our dexterous manipulation planning, we try to find intermediate grasp states that connect between the initial grasp and the final grasp of the dexterous manipulation, shown in figure 1.4. We do this by first segmenting a sequence of human demonstration into smaller subsequences using a segmentation criteria, which will be mentioned later on in subsection 1.3.2. Then, we will define our intermediate grasp states as the points which connect between these subsequences. These intermediate grasp states are very important information. They would be used as a clue for further extraction of the necessary skill parameters for the reproduction in the target system. The detail information of how they would be utilized are out of the scope of this thesis, and it would be mention later in Chapter 5 as a future work.

Since these intermediate grasp states can be considered as one type of the grasps because an object is still being hold in the hand, one might argue that this planning has no different from a continuous grasp recognition system which we used earlier to segment a stream of data to create a task model. However, this is not true, and we would like to emphasize a difference here.

As for a continuous grasp recognition, it first try to recognize pre-defined grasps and then segment the data into subsequences. On the other hand, our dexterous manipulation planning segment a sequence of data, in order to find those intermediate grasp states. The difference that we would like to mention, is not an order on which is done first between recognition or segmentation, but it is something that these two discovered. A continuous grasp recognition are build to discover type of grasps which are predefined. On the other hand, our dexterous manipulation planning are build to discover intermediate grasp states. Although these are also some types of grasps, they are *not predefined* grasps.

Number of Transitions and System Completion

In our system, we define a finite number of grasps that can be employed by a hand to grasp a particular object or a particular category of similar objects. For example, for object A or a category of objects similar to object A, only n type of grasps can be employed at the hand; Grp A1, Grp A2, Grp A3, ..., Grp An. Theoretically, a total number of transition between these grasps and the empty grasp is ${}_{n+1}P_2$ (indirect grasp). It seem that this number is too much when you want to teach the robot to dexterously manipulate among them until the system is completed. However, this is not true because in normal everyday life, some transitions are never occurred. Therefore, to teach a robot until it can do all dexterous manipulation for a particular object or a particular category of similar objects, we only need to teach all necessary transitions between all grasps. One of possible example is as shown figure 1.6, when we assume 8 types of grasps for Object A.

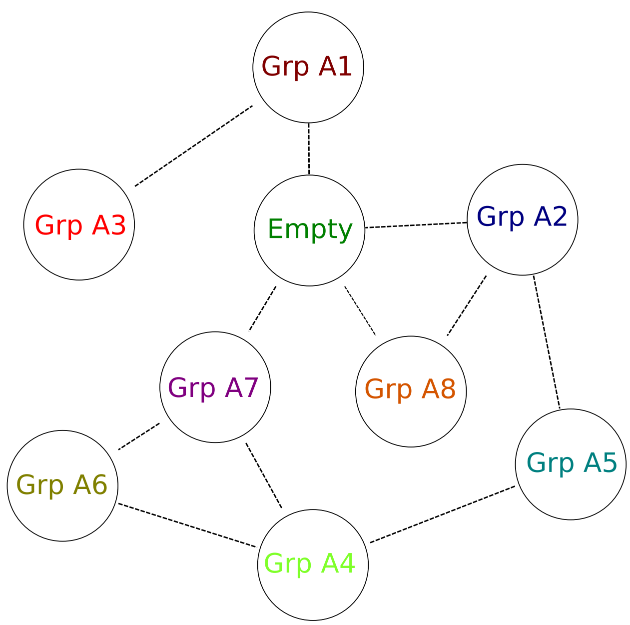


Figure 1.6: One possible example of grasp transitions graph for Object A

1.3.2 Segmentation's Criteria

Segmentation mentioned in the section is about a segmentation of the sequence of hand postures demonstrating a dexterous manipulation. The segmentation is conducted in order to extract or detect intermediate grasp states. An ideal criteria for this segmentation would be a detection of the change of contact relation between a grasping hand and the object. If the change and the position of all contact points can be recovered, we can then represent a planning of dexterous manipulation easily with these information. However, there are a few limitations on current technologies of a tactile sensors, which don't allow us to measure these information. One of the major limitation would be a size of each sensors, which leads to a discrete property of overall tactile sensor system. This is very important factor because we do not know exactly how dense the tactile sensors should be, in order to recognize a change of contact relation for grasping a particular object. Moreover, the more dense the system of tactile sensors are, the more sensitive and complicate of the recognition system would turn out to be. As a result, in our system we propose a method to detect the change of contact relation by considering a movement of finger of the manipulative hand.

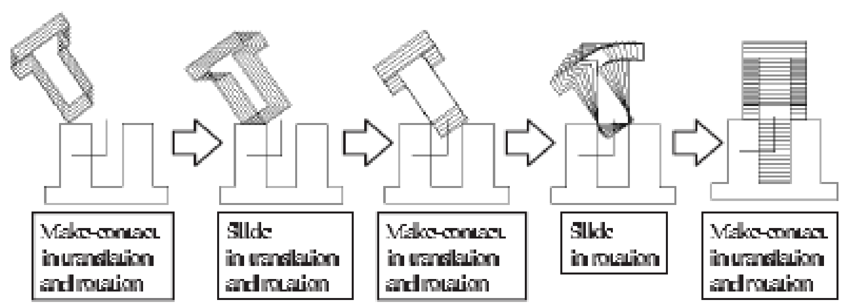


Figure 1.7: Transition of contact relation in assembly task; Takamatsu [1]

Takamatsu [1] has proposed a technique to recognize an assembly task from observation. In his work, he models his system based on contact relation of the objects. Although there are unlimited number of possible contact relations in an assembly task as you can imagine, he manages to categorize both a characteristic of contact relations and the transitions among them in to a limited number. What we are interesting in their work is the transitions among their category of contact relations. Although there are quite a few number of types of the transition, they are all based on two main characteristic: transition and rotation of the object as shown in figure 1.7.

In our study, if we look through the sequence of hand postures during dexterous manipulation, we can notice that at the moment where a major change or major transition of contact relation initially occurs, all joint angles of the grasping hand are also changing the way how they are moving together. In other word, we are saying that when all joint angles of the hand are moving coordinately in some pattern and suddenly change the pattern, it is a signal that the hand is about to change its contact relation with the object.

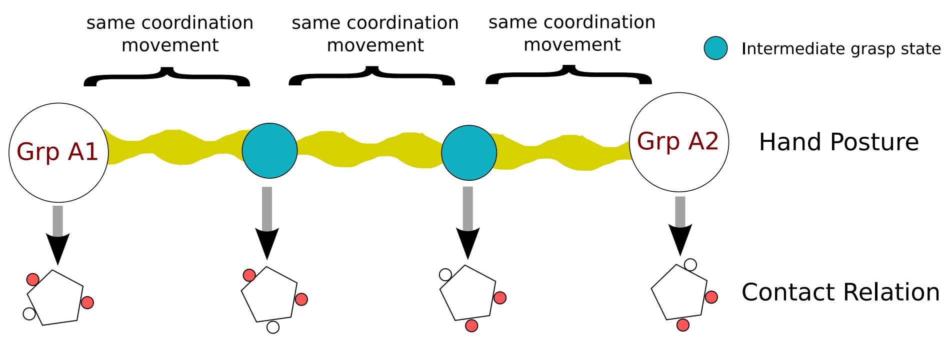


Figure 1.8: Criteria for segmenting a stream of hand postures demonstrating a dexterous manipulation

In our system, we segment the stream of dexterous manipulation based on the assumption given above. That is, we try to detect the points where the change of the coordinative movement of the hand occurs, and segment the stream at those points. The segmented point then will become our intermediate grasp states, which will later be used as the points where their contact relations considered to be important for dexterous manipulation planning.

A highly coordinated manner of joint angles of the hand during grasps has been validated in many previous studies [16][17]. At this point, it may look unclear what does it mean by saying that there is the coordinative movement of the joint angles of the hand, and also how to detect the change of them. As for a introductory of our technique, brief procedures are given below. More detail information could be found in chapter 3.

First, a dimension of joint angles of hand posture in the captured sequence will be reduced to three dimension using principle component analysis. The coordinate movement of the projected joint angles can be observed by an approximately linear relation in the new 3D joint space. Then, a curvature property will be used to detect the change of coordinative movement of a hand. The points where the change of coordinative movement occur are our intermediate grasp states.

1.4 Thesis Organization

In chapter 2 we introduce related works of the research. Firstly, we refer to a few researches to describe the characteristic of human hand. Then we move on to describe method currently used to extract intermediate grasp states of an dexterous manipulation.

In chapter 3, a detail definition of our system is given. Then our proposed method for intermediate grasp states extraction is explained. In chapter 4, three experiments have conducted to prove the efficiency of the proposed method. Another three approaches are also briefly introduced and conducted using the same data sets, in order to compare and evaluate the method.

Finally, in chapter 5 a summary of the thesis is given together with some ideas to improve and extend this research in the future.

CHAPTER 2

Related Works

In chapter 1, we have described various characteristics of our system, which we use for teaching a object manipulation task to a robot. There are many components in the system, which have to be implemented. However, as described earlier, we only consider a dexterous manipulation planning component, whose turn out to be an extraction of intermediate grasp states from stream of human demonstration.

In this chapter, a few previous studies on a characteristic of the hand posture during grasping are first reviewed. This should serve as a introduction and at the same time, give more detail information about human's hand. Then, previous studies about intermediate grasp states recognition are reviewed, based on two approaches: hand postures based approach and tactile sensors based approach.

2.1 Characteristic of Prehensile Movements

Grasp, or sometimes referred as prehensile movements, is defined as hand postures in which an object is seized and held within the hand. Many attempted has tried to explain and categorize these hand postures. However, there is not one that has been perfected. That is, there is always an exception for such explanations. Therefore, a few examples are given and reviews here, in order to give an overview of what has happened in this area of research. Some attempts has also considered a relation between a type of grasp and a manipulated object. We will try to convey them here as well.

Napier, J. R. [2]

This study characterized human prehensile movements into two categories: *power grasps* and *precision grasps*. Power grasp is a grasp which the object is hold inside a palm and opposition fingers, which exert a pressure it. It is utilized when stability and security a grasped object is required. On the other hand, if the object is held with the tips of the fingers and thumbs, where sensitivity and dexterity are dominated, it would referred as precision grasp. However, the study also stated that the two concepts are not mutually exclusive in some prehensile activities.

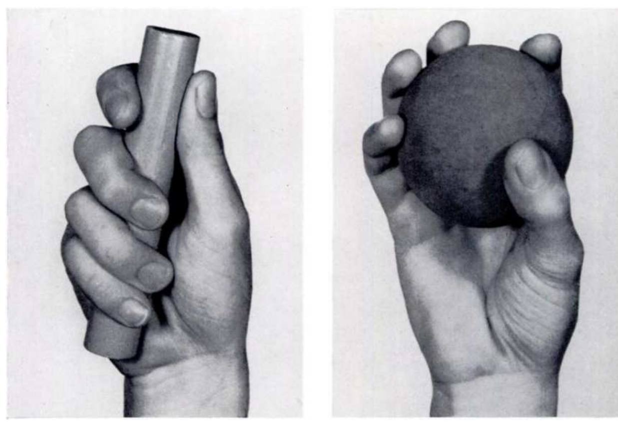


Figure 2.1: Power grasp (l) and Precision grasp (r); Napier[2]

Factors which influence the postures of the hand are also proposed in this study. It suggested that certain type of physical factors such as the weight, the texture, the temperature of the object may influence the type of grasp employed. However, in normal circumstance where any type of grasps can be employed, which postures would be adopted will depends solely upon the purpose to which the object to be hold for. In other words, intended activity has great influence on which type of grasps would be employed.

Kamakura et al. [8]

This study classified static prehension of hands into 14 patterns of prehension under four categories (Please refer to original work for pictures). The classification is based on position of fingers and contact area at the fingers. The underlying concept is that common patterns exist in finger use, so grasps should be able to classified regardless of the activities or objects.

As for a choice of grasp, the study believed that most of the time human would choose a type of grasp based on the specific group of objects they

Category	Class	Notation
Power Grips	Power Grip-Standard Type	PoS
	Power Grip-Hook Type	PoH
	Power Grip-Index Extension Type	PoI
	Power Grip-Extension Type	PoE
	Power Grip-Distal Type	PoD
Mid-Power-Precision Grips	Lateral Grip	Lat
	Tripod Grip-Standard Type	Tpd
	Tripod Grip-Variation I	TVI
	Tripod Grip-Variation II	TVII
Precision Grips	Parallel Mild Flexion Grip	PMF
	Circular Mild Flexion Grip	CMF
	Tip Grip	Tip
	Parallel Extension Grip	PE
Thumbless Grips	Adduction Grip	Add

Table 2.1: Kamakura's Grasp Taxonomy [8]

about to hold. This suggest that for specific type of objects, human have particular type of grasp for it.

Cutkosky, M. R. [3]

This study focused on how to select a suitable grasp for robot multi-fingered hand in manufacturing tasks. Cutkosky believed that by given an initial information about the task requirements and object shape, a question about choosing a good grasp might able to be resolved. An expert system had been constructed based on the grasp taxonomy. While having a feedback information from machinists, the system will travel along the tree of the taxonomy and made a decision for a grasp.

After some revisions of the system, the study suggested that a prediction on how people would grasp parts and tools in particular environment can be done. However, the system itself is far from complete, and may never complete. The study also suggested that the taxonomy has been useful, but it has limitations. One of which is that, it is incomplete. An effective way to extend it, is to considered a concept of *virtual finger*[16]

The study also discussed about the problem of choosing a grasp based on analytic approaches. Cutkosky suggested that when an experiment have to been done in uncontrolled environment, the approach may have a problem of measure and modeling some properties of a grasp.

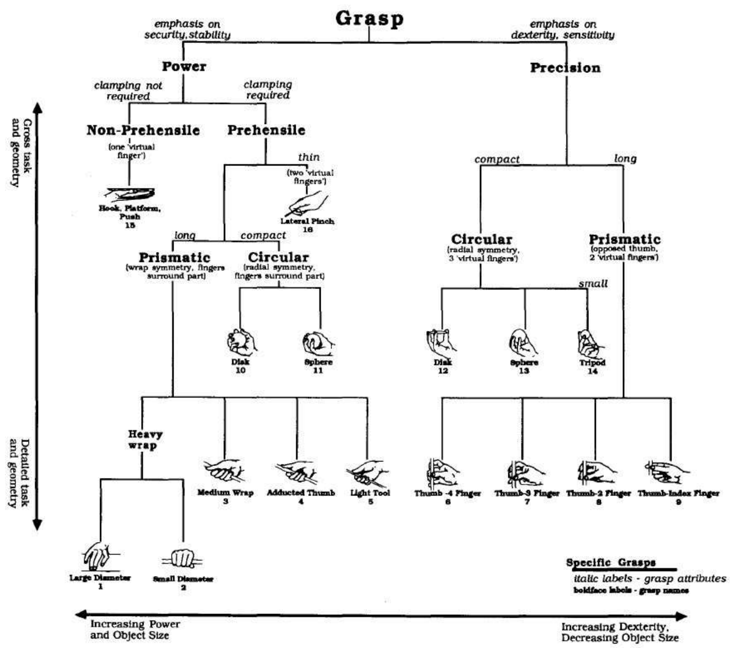


Figure 2.2: Grasp taxonomy used by Cutkosky's expert system[3].

Arbib, M. A. and Iberall, T. et al. [16][18]

Arbib et al.[16] introduces a concept of *virtual finger*. Virtual finger is an abstract representation, which combine physical fingers or hand surfaces into a unit. One unit of virtual finger represents those individual fingers or hand surfaces that work together as a unit and apply same direction of force or torque at the object. This virtual finger insists a belief that there is high correlation of among fingers during human's static prehensile movement.

Iberall et al.[18] proposed a concept of opposition space. Opposition force refer to forces that oppose to each other at the object during human's prehensile movement. These forces can be created by either virtual finger or the palm. Iberall suggests that there are three type of opposition forces, which are *pad* opposition, *palm* opposition, and *side* opposition. These oppositions are primitives in the opposition space. Human's prehensile postures will be defined in opposition space as a combination of these opposition primitives.

Santello et al. [17]

In this study, human hand postures during grasp were analysed statically. Each hand postures is represented with 15 joint angles measured by data glove. Subjects were asked to shape their hand as they are grasping an object.

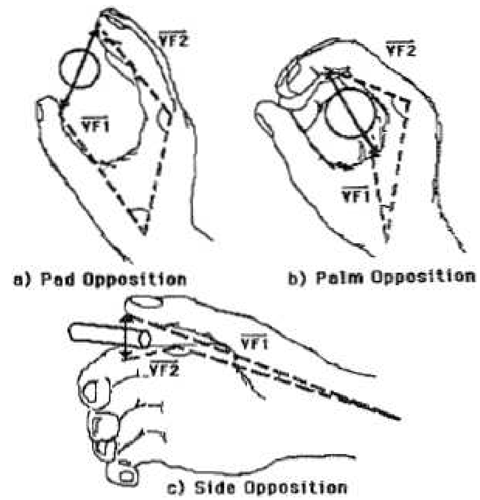


Figure 2.3: Virtual finger and three basic oppositions (figure [4])

Five subject, and 57 test objects were used. Three techniques in static were used for analysing the data; discriminant analysis regression analysis, and principal components analysis (PCA).

PCA indicated that 80 percents of variances of the data could be covered by the first two or three principle components. In other words, 15 dimensions of joint angle which originally represents hand postures, can reasonably be reduced to two-three dimensions, with a small amount of loss data.

Result of this study revealed that for static hand postures grasping various objects, all joint angles are not controlled independently, but they are rather have high correlation among them. In the same manner, it might could be believed that static prehensile movement of human are not classified into discrete representation as other grasp taxonomy, but rather a continuous representation in this principle analysis space.

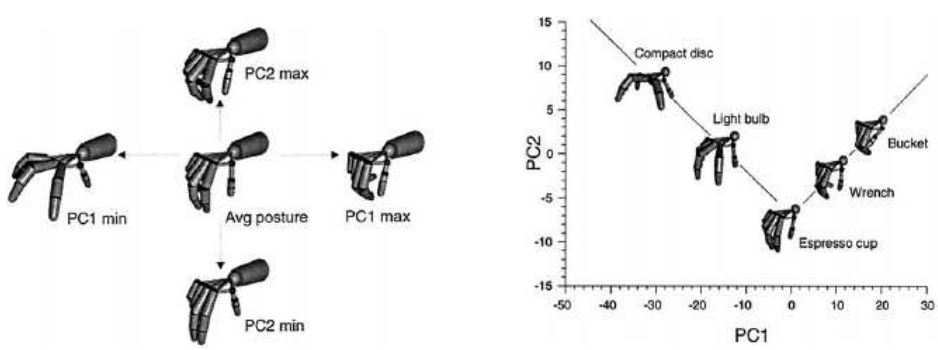


Figure 2.4: Hand postured in reduced joint angle space (2D)

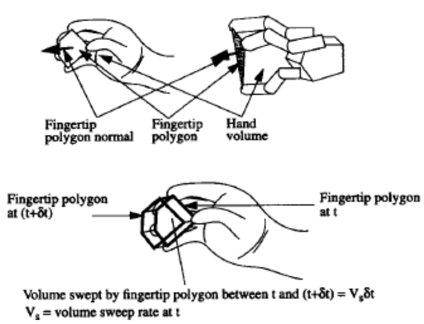
2.2 Intermediate Grasp State Extraction

In this section, we review some previous researches which have been proposed to segment a demonstration stream of dexterous manipulation. Although there are so many researches about segmenting and recognizing grasp from human demonstration, but only a few are about segmenting a dexterous manipulation. The segmentation approach can be divided into two groups; hand postures based approach, and tactile based approach. We will review them separately. Note the some techniques mention here are also from researched proposed to solve grasp segmentation problem, but the techniques might be of used to understand the less of the techniques.

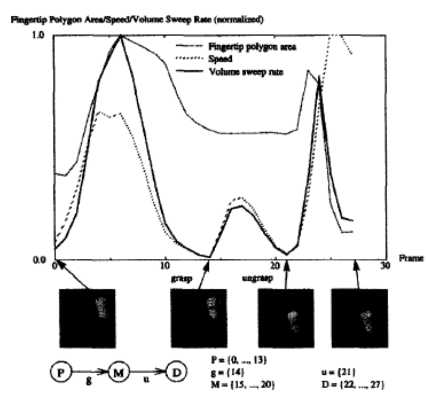
2.2.1 Hand Posture Based Approach

Kang, S.B. and Ikeuchi, K. [5]

Kang [5] proposed a technique to segment human hand motion in object manipulation. In the study, an object manipulation are composed of many subtasks, where each subtasks is combination of pre-grasp phase, grasp phase, manipulation phase, place object phase, and depart phase sequentially.



(a) Features used in segmentation



(b) Example of experimental data

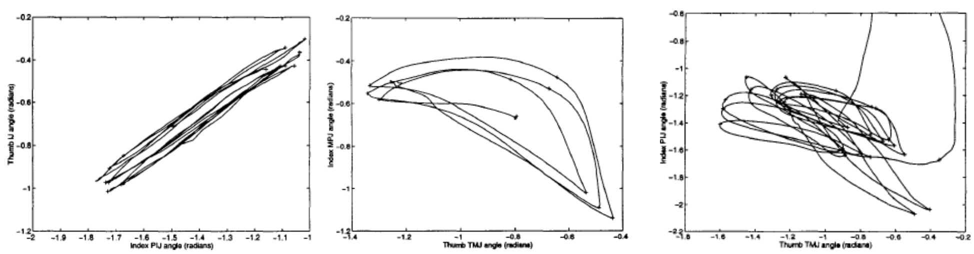
Figure 2.5: Temporal segmentation for object manipulation; Kang [5]

To segment each subtasks, an analysis of the three main features are used, which are the fingertip polygon area, the speed of hand movement, and volume sweep rate. Fingertip polygon area can be easily calculated when the position of fingertips (except thumb) are known. A speed of hand movement is captured using motion capture system, and volume sweep rate is a product of fingertip polygon area and volume sweep rate.

Zacksenhouse, M. and Moestl, T. [19]

Zacksenhouse [19] proposed a technique to segment dexterous manipulation movements. In the study, human dexterous manipulation movements were classified as either simultaneous or sequential[7]. Simultaneous movements are movements where all participating joints move in the same pattern simultaneously. Sequential movements are movements which is a combination of multiple simultaneous movements.

Two most active joints defines phase-plane for each movements. In the phase-plane, it is believed that the trajectories of simultaneous movements would be approximately linear. Their segmentation technique of sequential movements is based on this feature.



(a) Sequential Movements (b) Complex Manipulation

Figure 2.6: Experimental results of Zacksenhouse's segmentation technique

The experimental result shows that the technique work well in sequential movements, which are a multiple of same simultaneous movements. An example of such movements are screw & unscrew, close a cap etc. However, when it came to more complex manipulation, which in this experimental is cap pinching movement followed with cap closing movement, the technique cannot segment the movement effectively. This is because during complex manipulation, many more joints may participate in the movement, and only single phase-plane may not enough to analyse this. Zacksenhouse had suggested that multiple phase-plane may be of used in this case.

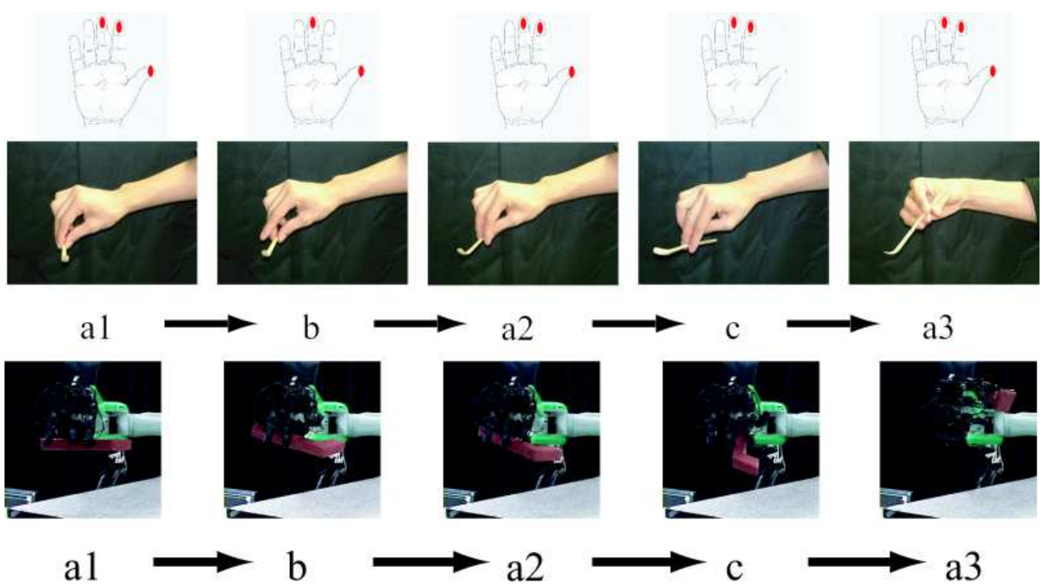
2.2.2 Tactile Sensor Based Approach

Kudoh et al.[6]

In this research, a whole framework for learning everyday object manipulation from observation is proposed. Dexterous manipulation phase is also considered. In the study, possible dexterous manipulation between 14 Kamakura grasp taxonomy [8] are segmented and planned.

Data glove with 13 tactile sensors is designed in order to record human demonstration. A segmentation technique is performed based on the belief that only one contact point changes in between one step of transition of dexterous manipulation. Although the results are mapped and reproduced in target robot, a segmentation process is still lack of automation. The study has shown some result captured from 13 tactile sensors, but it has not proposed a effective method to measure change whether the particular sensor is in contact with object or not.

(a) 13 tactile sensors configuration



(b) Segmentation and reproduction result from PMF grasp to Tpd grasp

Figure 2.7: Data acquisition and experimental result; Kudoh[6]

CHAPTER 3

Intermediate Grasp States Extraction

In this chapter, we explain about the method we use to segment a stream of dexterous manipulation. The method is explained in details together with the reason why it should be effective. Principle component analysis and curvature are two fundamental techniques that have been utilized in the proposed method.

Before we can get to method's details, the formation of the system is given below. It specified the overall system, including our grasp taxonomy, data acquisition system etc.

3.1 System Formation

Object manipulation tasks we are interested in this study are those tasks in a human painting activity. During painting, a lot of objects and tools are used. This leads to a result that there are a lot of object manipulation occurred, e.g. grasping, releasing. Moreover, since for one particular object there are many types of grasp employed depending on the purpose of painter, a lot of transitions between those grasps (or more formally, dexterous manipulation) are also taking place during the painting process.

To be more specific, in this study we focus on dexterous manipulation of the paint brush or some other tools with similar usability e.g. pen, pencil. Object's geometry, e.g. size or weight, seems to effect how human would grasp and manipulate the object [2]. However, these factors will not be taking into consideration. In this study, we consider only the paint brush which size and weight are reasonably relative to demonstrator's hand, and would not effect demonstrator's ordinary manipulation movement.

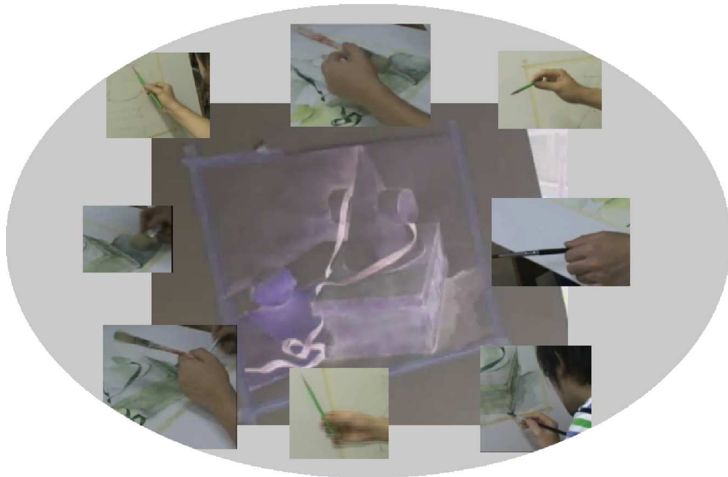


Figure 3.1: Object manipulation during painting activity.

One interesting issue is that paint brush used in demonstration and in reproduction with robot hand are also different in geometry. This may seem peculiar, but there is also a supporting reason. In planning, the output we are expecting is a kinematics property of a hand, not a dynamic one. Moreover, in real robotic hand, we also assume that the hand should be designed in order to cope with average forces and pressures in its environment.

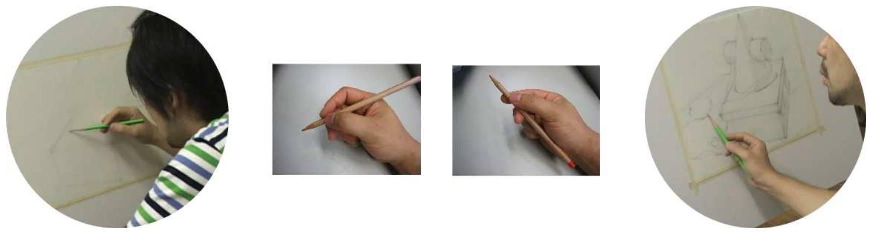
In addition, according to Napier [2], grasps are also affected by other physical factors like the texture, the temperature, wetness of the object etc. It is true that in this study we would like to plan a dexterous manipulation planning to imitate human movement and intention, but as for preliminary step, those extreme cases are considered.

3.1.1 Defining Grasp Taxonomy

Grasp taxonomy defined in this study does not follow any other previous studies. Since we are interesting in grasps and dexterous manipulation of a paint brush in human painting activities, we went to observe and ask professional artists directly. There are six types of grasp in our taxonomy. They are only grasps which used with paint brush or some similar tools. This is done based on previous study [8], which suggests that human do have a specific type of a specific category of hand posture to grasp some specific tools or objects. Our taxonomy is divided and shown in figure 3.2-3.4 based on their purpose of use.

Drawing Grasps

There are mainly two types of grasp for drawing. The main different is that type 1, shown in figure 3.2 on the left, uses the movement wrist and fingers to draw, but on the other hand type 2 uses the movement of whole arm in order to draw.

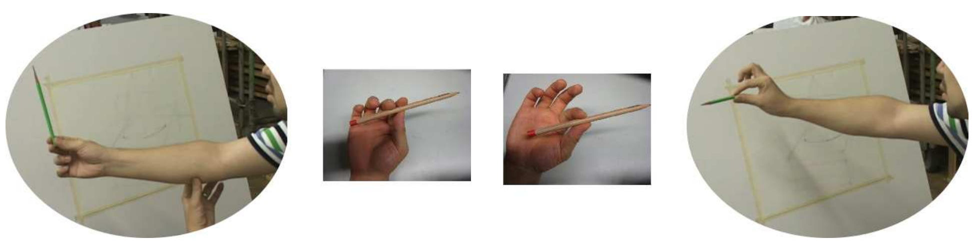


(a) Drawing grasp type 1; D1 (b) Drawing grasp type 2; D2

Figure 3.2: Grasp used for drawing and painting

Measuring Grasps

In measuring, most artists use two types of grasp. Type 1, shown on the left of figure 3.3, are hold tightly by three fingers. This is because a stability is needed. On the contrary, only two fingers are used to hold type 2 in order to enable flexibility, which sometimes needed when the artist wants to use this type of grasp to draft a slightly thin line.

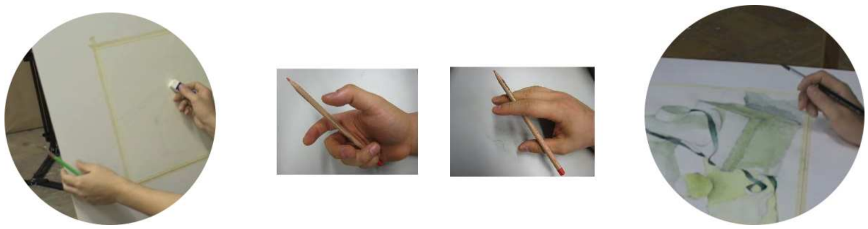


(a) Measuring grasp type 1; M1 (b) Measuring grasp type 2; M2

Figure 3.3: Grasps used for measuring

Other Grasps

Apart from drawing and measuring grasps, artists also have another two types of grasp for a paint brush; holding and resting. The objective of these two grasps are very similar. While holding grasp is employed by either hands in order to enable another hand or finger of holding hand to be available for other tasks, the purpose of resting grasp is, just like its name implied, only for resting a hand.



(a) Holding grasp; H1 (b) Resting grasp; R1

Figure 3.4: Grasp used for other purposes

3.1.2 Data Acquisition System

In our system, a combination of two system are used to capture human demonstration.

- *CyberGlove*: A 18-sensor data glove provides information about hand posture. 18 joint angles of the hand is measured and fed back in real time. We can represent this using 18 dimensions vector.
- *Polhemus FASTRAK*: An industry standard motion tracking system offers six degree of freedom; three for position and three for orientation, in real-time. In the system, three receiver sensors are used to provide information about hand and object location.

A combination of the two system provides us a data up to the resolution of 110 frames per second in real-time. A system setup is shown in figure 3.5

3.1.3 Task Primitives Segmentation & Recognition

There are many object manipulation (grasping, manipulating, releasing) occurred during painting activities. Many kind of objects were used, for instance a paint brush or a pencil, a ruler, a rubber etc. In section 1.3.1, we have already explained that in order to build a task model, pre-defined task primitives has to be segmented and recognized out of a stream of human demonstration.

Since our main interest is not task primitive segmentation or recognition, we decide to do this manually. In other words, in actual system we demonstrate and record each transitions between each pair of grasps separately.

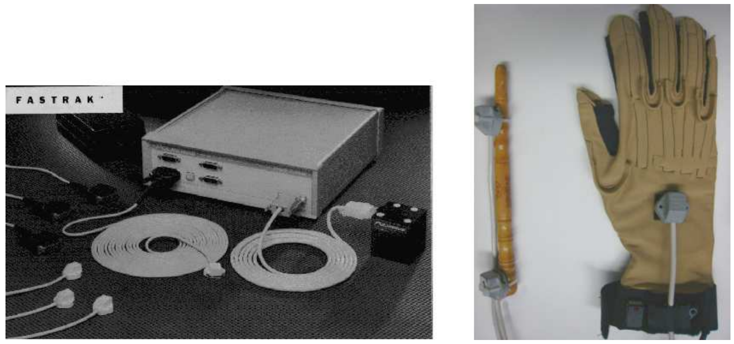


Figure 3.5: System setup of a data acquisition system

By doing all of these manually, we have raised another important problem. We could not know which translation are necessary in real life, which would prohibit us from creating a graph of grasp transition as shown in figure 1.6. However, this is also out of the scope of this thesis.

Although we have not constructed a segmentation system, we have built a static grasp recognition. This is done with a purpose to identify type of grasps, together with the type of transitions.

Static Grasp Recognition

The grasp recognition system is constructed. Six types of grasp in figure 3.2-3.3 are classified using support vector machine. Each hand postures are represented with 18 dimension vector, where each of which is corresponded to each joint angles of the hand. Five data sets are captured. Each of which contains around 290 sample data. Each one are used as training data and testing data exchangeable, and experimental results are shown in figure 3.6.

Note that a recognition is also tested on three dimensions data of hand postures as well. They are a reduced version of original 18 dimensions of joint angles using principle component analysis. For more detail about how to reduce the dimension of hand postures, please refer to section 3.2.1.

3.2 The Proposed Method

In this section, we explain about how we extract intermediate grasp states from a sequence of hand postures of a dexterous manipulation. Before going into more details about the techniques proposed, we would like to start by defining the problem more precisely.

In each frames of data captured from data acquisition system, it composes

	Test0 (280)	Test1 (310)	Test2 (293)	Test3 (293)	Test4 (292)
Test0	3D = 99.64 18D = 100	3D = 88.06 18D = 97.10	3D = 99.32 18D = 98.29	3D = 92.15 18D = 95.22	3D = 96.58 18D = 100
Test1	3D = 89.58 18D = 94.58	3D = 99.63 18D = 100	3D = 99.60 18D = 98.40	3D = 89.68 18D = 91.27	3D = 99.60 18D = 100
Test2	3D = 85.83 18D = 92.92	3D = 79.33 18D = 95.20	3D = 100 18D = 100	3D = 100 18D = 100	3D = 100 18D = 100
Test3	3D = 90.83 18D = 92.92	3D = 86.36 18D = 90.77	3D = 100 18D = 100	3D = 100 18D = 100	3D = 100 18D = 100
Test4	3D = 85.00 18D = 94.58	3D = 84.50 18D = 92.99	3D = 100 18D = 100	3D = 97.62 18D = 100	3D = 100 18D = 100

Figure 3.6: Experimental results of static grasp recognition using SVM

of two part; one from data glove, and the other from motion tracking system. In this section, we only consider the data from data glove, which represents joint angles of the hand posture. Our data glove can measure up to 18 joint angles of the hand. Therefore, any hand posture can be represented as

$$\mathbf{p} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \dots & \theta_{18} \end{bmatrix} \in \mathbb{R}^{18} \tag{3.1}$$

,which is a 18 dimensional vector in 18 dimensional joint space. Then, a stream of dexterous manipulation demonstration can be written as a sequence of hand postures as

$$\mathbf{S} = \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_T \tag{3.2}$$

where $\mathbf{p}_t; t \in 1, \dots, T$ refers to a hand posture at time t and T is the number of frames in the sequence.

Therefore, we can now define our intermediate grasp states extraction as a problem of finding hand postures $\mathbf{p}_{i1}, \dots, \mathbf{p}_{iN}; i1, \dots, iN \in 1, \dots, T$ that divide sequence $\mathbf{S}$ into subsequences, where each subsequences represents a coordinative movement of the hand.

After the problem is clearly describe, we can now start off with the proposed method. Since there are many procedures in our proposed method, we first give the overview of it below, and the detail explanation of each procedures in the following subsection.

1. Reducing the 18 dimensional vector of the hand posture to three dimensional vector using principle component analysis (PCA).

   (a) Forming new 3D coordinate space using the first three principle components (eigenvectors) as basis.

(b) Projecting a sequence of 18 dimensional vector of hand postures into newly created 3D coordinate space, which will result in reducing the dimension of original data to three dimension. This means that the input sequence of 18 dimensional vector of hand postures has turned to be a sequence of three dimensional vector of data.

2. Calculate curvature of a sequence of three dimensional vector of data, and detect intermediate grasp states as the local maximum in curvature-time graph. This is done based on our assumption that if we project a sequence of hand posture which represents a coordinative movement of the hand into a reduced three dimensional joint space creating by PCA, it would result in an *approximately linear* trajectory of a sequence of projected hand postures.

3.2.1 Dimensional Reduction using PCA

Principle component analysis has widely used as a technique to reduce a dimension of highly dimensional data. It has just been first introduced and applied to highly dimensional data of joint angles of hand posture in the last decade [17]. Recently the technique has been applied in the research field of grasp planning and referred as *eigengrasp* [20]. In this subsection, we first revise a technique called eigengrasp, and then explain how we apply this technique as part of our proposed method.

For all d joint angle hand postures $\mathbf{p}_i = \begin{bmatrix}\theta_1 & \dots & \theta_d\end{bmatrix}$ and N is number of hand postures, we calculate their principle components [21] by

1. Calculate average (mean) hand posture $\bar{\mathbf{p}} = \begin{bmatrix}\bar{\theta_1} & \dots & \bar{\theta_d}\end{bmatrix}$, and subtract each hand postures by the mean; $\mathbf{p} - \bar{\mathbf{p}}$

2. Calculate the covariance matrix $\mathbf{C}^{d*d}$, when consider each joint angles as one dimension. As the name implied, each elements of the matrix are represented with the covariance of two joint angles, $\mathbf{C}_{ij} = cov(\theta_i, \theta_j)$

3. Calculate eigenvectors and eigenvalues of the covariance matrix, then rearrange the order of eigenvectors based on their corresponding eigenvalues from high to low. We refer to them as $\mathbf{e}_1, \dots, \mathbf{e}_d$ and $\lambda_1, \dots, \lambda_d$, where $(\mathbf{e}_i, \lambda_i)$ pair is a d-dimensional eigenvector and its corresponding scalar eigenvalue. This is where the name *eigengrasp* comes from, when eigengrasps are referred to these eigenvectors which are the principle components of input hand postures.

The technique to calculate eigengrasp is described above. There might be some minor difference on what kind of input hand postures are used, which sometimes effects the fraction of variance of input hand postures accounted by each eigengrasps. A fraction of variance accounted by eigengrasp $\mathbf{e}_i$ is given by

$$\tilde{\lambda_i} = \lambda_i / \sum_{k=1}^{d} \lambda_k \tag{3.3}$$

, and accordingly a variance accounted by the first b eigengrasp is given as $\sum_{k=1}^{b} \tilde{\lambda_k}$.

Santello [17] used a huge amount of static hand postures, when subjects were asked to shape their hands as they were grasping various kind of objects. His result shows that the first two eigengrasps could account for more than 80% of the variance. On the other hand, Dejmal [22] used all hand postures in multiple cycles of a simultaneous hand movement as an input to create eigengrasp. Her result suggests that the first eigengrasp accounts around $91.7 \pm 7.0\%$ of the variance.

In our proposed method, we use all hand postures in a stream of dexterous manipulation demonstration, $\mathbf{p}_1, \ldots, \mathbf{p}_T$. This is different from Dejmal [22] that in our stream of dexterous manipulation, it is not only the multiple cycles of one simultaneous hand movement, but it may compose of many kind of hand movement included simultaneous hand movement. From our experiment, which three different kind of dexterous manipulation demonstrations are considered, it had shown that on average the first three eigengrasps account at least 90% of the variance of original hand posture in the input stream.

As for the reason given above, we decide to reduce the dimension of our input hand postures to three dimensions. This can be done by

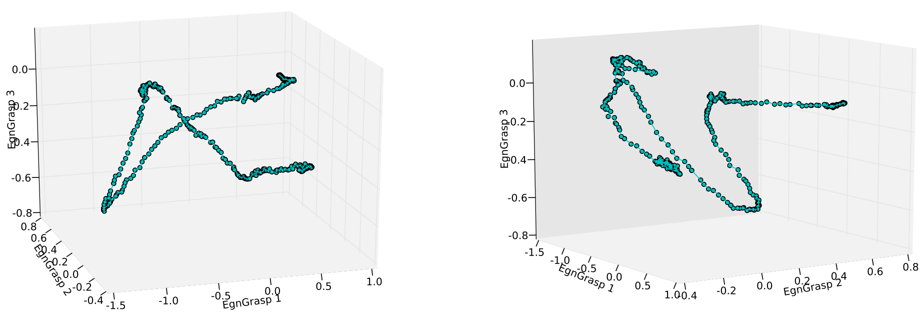
1. Forming a new coordinate space using the first three eigengrasps as basis. For the sake of simplicity, we will refer to this as *3D-eigengrasp space.*

2. Projecting original mean-adjusted hand postures, $\mathbf{p}_i - \bar{\mathbf{p}}$, to 3D-eigengrasp space. This results in the representation of all original hand postures in this subspace, which becomes

$$\mathbf{p}_i = \sum_{k=1}^{3} y_{i,k} \mathbf{e}_k \tag{3.4}$$

, where $y_{i,k}$ is an amplitudes of the hand posture $\mathbf{p}_i$ along eigengrasp $\mathbf{e}_k$. In other words, we can express each dimensional-reduced hand postures as three dimensional vector of the amplitudes,

$$\mathbf{y}_i = \begin{bmatrix} y_{i,1} & y_{i,2} & y_{i,3} \end{bmatrix} \in \mathbb{R}^3 \tag{3.5}$$



(a) View 1 (b) View 2

Figure 3.7: Example of trajectory of hand postures in 3D-eigengrasp space

An example of trajectory of sequence of projected hand postures is shown in figure 3.7. Axes of the graph are $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ respectively.

3.2.2 Extract Grasp States using Curvature

Consider carefully, it might be able to notice that in figure 3.7 the trajectory is composed of many linear portions connecting together. They may not be a straight line, but we can approximate them roughly as linear relation. This assumption supports and gives up some hint on how to extract intermediate grasp states.

As explained in section 1.3.2, our intermediate grasp states are the moments in the stream of dexterous manipulation, where the change of the coordinative movement of the hand and also the change of contact relation between the hand and object initially occur. Note that, by saying *coordinative movements of the hand*, we refer to a period that all joint angles of the hand are moving together in the similar pattern at the whole period of time.

If we look through experimental result in Dejmal [22] carefully, we may notice that their simultaneous hand movements are also some kind of coordinative movements of the hand. Furthermore, each of those simultaneous hand movements can also be considered as a line in 3D space, since 90% of hand postures in the movement can be accounted by its first eigengrasp.

Although in our system we don't consider only those simultaneous hand movements [7], we may generalize from Dejmal's experimental result that each of coordinative movements of the hand are also behaved as a line or some kind of linear relation in 3D-eigengrasp space. Therefore, in order to extract intermediate grasp states of the dexterous manipulation, we have to find the points of trajectory of the dexterous manipulation in 3D-eigengrasp space where two continuously linear lines intersect.

Although it may seem similar to straight lines connecting together in figure 3.7 by human judgement, it is very difficult for computer to interpret where they intersect. Therefore, instead of trying to map those linear-liked sub-trajectories to lines and find there intersects, we decide to consider a curvature of the trajectory. A curvature is the amount which imply how much the trajectory or curve is deviated from being straight. In our system, we would like to find the points where lines are intersect, or in other word, where the corners occur because the trajectory or curve is changing its direction. This can also be interpreted as points where their curvature are high. To sum up, in the context of curvature, our intermediate grasp states are those points in the trajectory where their curvature is higher then some given threshold, and it would be considered as a local maximum if we plot the graph of curvature of every points of the trajectory.

For parametric curve $\mathbf{r}(t) = (y_1(t), y_2(t), y_3(t)) \in \mathbb{R}^3$, curvature $k(t)$ is defined as

$$k = \frac{d}{ds}\mathbf{T} = \frac{d\mathbf{T}/dt}{ds/dt} = \frac{\dot{\mathbf{T}}}{|\dot{\mathbf{r}}|} \tag{3.6}$$

, where $\mathbf{T} = \dot{\mathbf{r}}/|\dot{\mathbf{r}}|$ is unit tangent vector. Curvature is the rate of change of $\mathbf{T}$ with respect to arc length. It tells us how fast the unit tangent vector is changing and in which direction. In our system, we are only interested in how fast the unit tangent vector is turning, which is the magnitude of the curvature given above. Moreover, our data is discrete and the sampling rates between each pair of consecutive hand postures are also equals, so we can simplify the calculation of curvature.

Given three time-consecutive projected hand postures as points in 3D-eigengrasp space $\mathbf{y}_{t-1}$, $\mathbf{y}_t$, and $\mathbf{y}_{t+1}$: a magnitude of the curvature at time t is defined as the size of an outer angle in degree between vector $\overrightarrow{\mathbf{y}_{t+1}\mathbf{y}_t}$ and vector $\overrightarrow{\mathbf{y}_t\mathbf{y}_{t-1}}$. The geometric interpretation of this is shown in figure 3.8, and the mathematic representation is written as

$$|k(t)| = \arccos \frac{\overrightarrow{\mathbf{y}_{t+1}\mathbf{y}_t} \cdot \overrightarrow{\mathbf{y}_t\mathbf{y}_{t-1}}}{|\overrightarrow{\mathbf{y}_{t+1}\mathbf{y}_t}||\overrightarrow{\mathbf{y}_t\mathbf{y}_{t-1}}|} \tag{3.7}$$

. By calculating the magnitude of curvature as shown in figure 3.8, we can easily apply the concept of *scale space* to handle noisy data as well. This

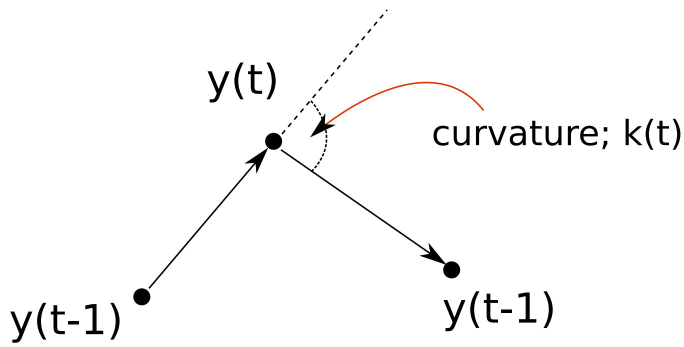


Figure 3.8: Simplified version of how to calculate curvature

can be done by considering points $\mathbf{y}_{t-s}$, $\mathbf{y}_t$, and $\mathbf{y}_{t+s}$ instead when calculate $|k(t)|$, and changing scale s until achieving a reasonable result.

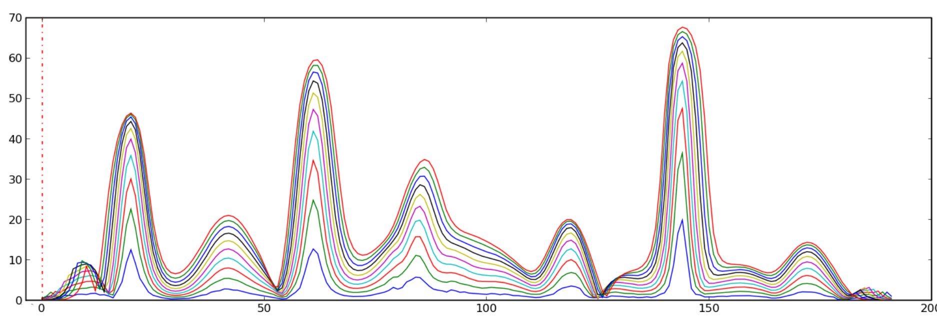


Figure 3.9: Curvature of trajectory in figure 3.7, in 10 different scales

An example of the curvature calculation of trajectory in figure 3.7 is given in figure 3.9. After some preprocessing with the trajectory, a curvature is calculated for $s = 1, \ldots, 10$.

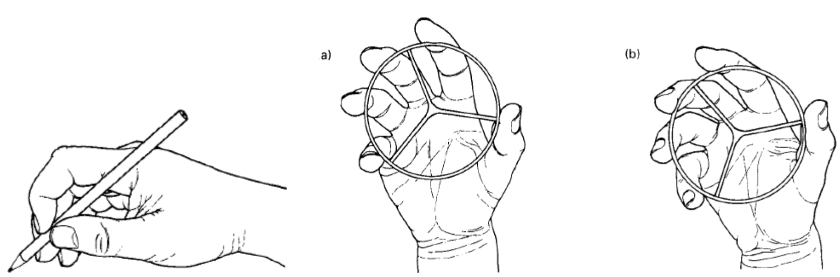
CHAPTER 4

Experiment and Evaluation

An experiment is done on three different streams of dexterous manipulation. Another three different methods which approaches similar problem are also briefly introduced. Results from all methods (experimental setup), including ours, are compared and evaluated. Then, some discussions is given at the end of the chapter.

4.1 Experimental Data and Setups

Three different movement of dexterous manipulation are considered as our experimental data. Each of them is classified as a movement that occurs in human usual manipulative movement[7]. However, how they are conducted may be different in person.



(a) Dynamic tripod (b) Rock

Figure 4.1: Sketch of dynamic tripod and rock movements; Elliott[7]

1. *Dynamic tripod*: In this experiment, a drawing grasp type 1 is employed at the pen, and the hand is moved as it is writing. Each data sets composes of this movement eight times back and forth.

2. *Rock*: In this experiment, both hands grasp a Rubic cube, and the right turns it counter clockwise five times in a row. Only the movement of the right hand is considered and captured.

3. *Interdigital step*: In this experiment, at first a drawing grasp type 1 is employed at the pen. Then the hand try to change type of grasp to a drawing grasp type 2. A Sketch of this movement is shown in figure 4.2, but in the real situation the step in between is a little different.

Note that in experiment 1 and 2, they do not follow the system formation given in section 3.1. In experiment 1, this type of movement will be considered only as a homogeneous manipulation, not a dexterous manipulation, because type of grasp is not changed at all (according to our painting grasp taxonomy). In experiment 2, both the object and type of grasp employed are both different and not defined in section 3.1. However, two of these experiment are conducted in order to compare our result with the other method which claim to be able to achieve them. Moreover, in experiment 2, we intentionally choose different object, because we also want to show that our proposed method can also be utilized with various objects.

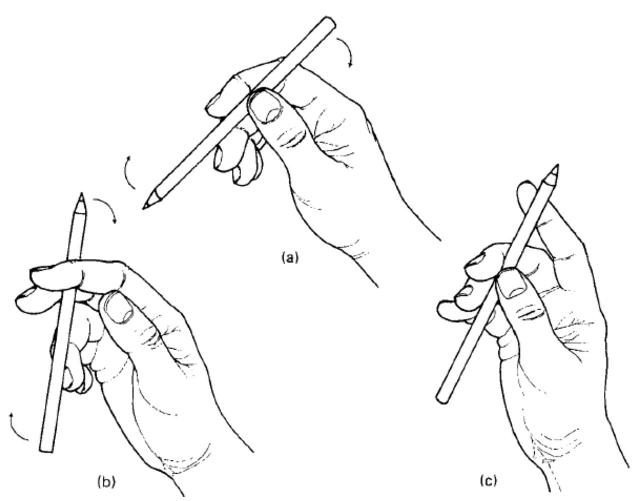


Figure 4.2: Sketch of interdigital step movement; Elliott[7]

To deal with inconsistency, we capture around 5-10 data sets of each movements. All are captured with the highest capacity of the acquisition system; 110 frames per second. Video is used for visualization of the data. However, the acquisition system and the video are not 100 percents synchronized, rather they are matched manually.

We conduct each experiments with the experimental setup given in subsection 3.2: reducing dimension of hand postures using PCA, and extracting intermediate grasp states using curvature. However, we add some variations here in the process before projecting hand postures to 3D-eigengrasp space, which we describe them below.

Exp. Setup 1

This experimental setup do not vary from the proposed method explained in section 3.2. It is all same. 3D-eigengrasp space is created using by taking all eighteen joint angles as input. Then, all input data are projected into the eigengrasp space, and the intermediate grasp states are extracted using curvature.

Exp. Setup 2

This experimental setup has a small variation from Exp. Setup 1. The different is that 3D-eigengrasp space is created using all data sets in the experiment. In other words, we project all hand postures in one movement, not it own principle component space, but to the principle component space of the movements of the similar kind. We know that this might seem awkward, because when we want to adjust the mean of our input trajectory, we may get confuse which one to use between the average of the input trajectory or the average of the trajectories that used to create 3D-eigengrasp space. However, we try this variation in order to see whether there would be any different from the variation, because using only one input data sets seem a little risky.

Exp. Setup 3

This experimental setup has the most variation from the proposed method explained in 3.2. All processes that have to be done are same, but the input are different. Instead of consider only eighteen joint angles as each frames of the input, we also include speed of each joint angles at that particular time.Therefore, each frames of input data becomes $\mathbf{p} = \begin{bmatrix}\theta_1 & \dots & \theta_{18} & \dot{\theta_1} & \dots & \dot{\theta_{18}}\end{bmatrix} \in \mathbb{R}^{36}$

Note that value of speed of the joint angles are very small, compared to the value of joint angles. Therefore, in the real situation when we calculate PCA of all the data, we also magnify the speed by some scales. Otherwise, the 3D-eigengrasp space will not have any different from the one in Exp. Setup 3.

The reason why we set up this experiment like this is that, we add one assumption to the characteristic of intermediate grasp states. That is, intermediate grasp states are those points where most joint angles and joint speeds do not changes their value, or in other words, a brief stop of a hand. This is the main reason why we add speed of the joint angles to our input.

Despite the fact that after the reduction of dimension of the combination of joint angles and joint speeds the meaning of principle components and curvature are not clearly understood, we consider local minimums as intermediate grasp states because they are points in the trajectory where joint angles and the speed of the joint angles does not change. (While local minimums are estimated as brief stop of hand , local maximums are very difficult to interpret. It could be understood as points where either joint angles or joint speeds change their value.)

In our experiment, although their are very few researches concentrating on the same topic, we also try to compare our result with others. Another variation of our experimental setup is given below, and it is an implementation that based no an assumption given by Zacksenhouse [19].

Exp. Setup 4
In this experimental setup, we consider two most active joint angles from eighteen joint angles. Zacksenhouse does not specify exactly how to do this, so we select them manually based on our observation. Then the second step is to extract intermediate grasp states. Originally, zacksenhouse to suggest that when consider a graph using the two most active joint angles as axis (called phrase-plane), coordinated movements would generate a trajectory that is piece-wise approximately linear. Therefore, she segments at points that break the trajectory into subtrajectory that are approximately linear. Since her assumption is very similer to how er extract intermediate grasp states from 3D-eigengrasp space, we also use curvature to find her segment points in trajectory in the phase-plane.

4.2 Experimental Results

In each experiments, only one out of all captured streams are used as the input data. The experimental result of our proposed method (Exp. Setup 1) is first shown in detail. Then, it is compared with the results from other experiments setup one by one. Finally, a conclusion is given for each experiments.

4.2.1 Experiment 1 : Dynamic Tripod of the Pen

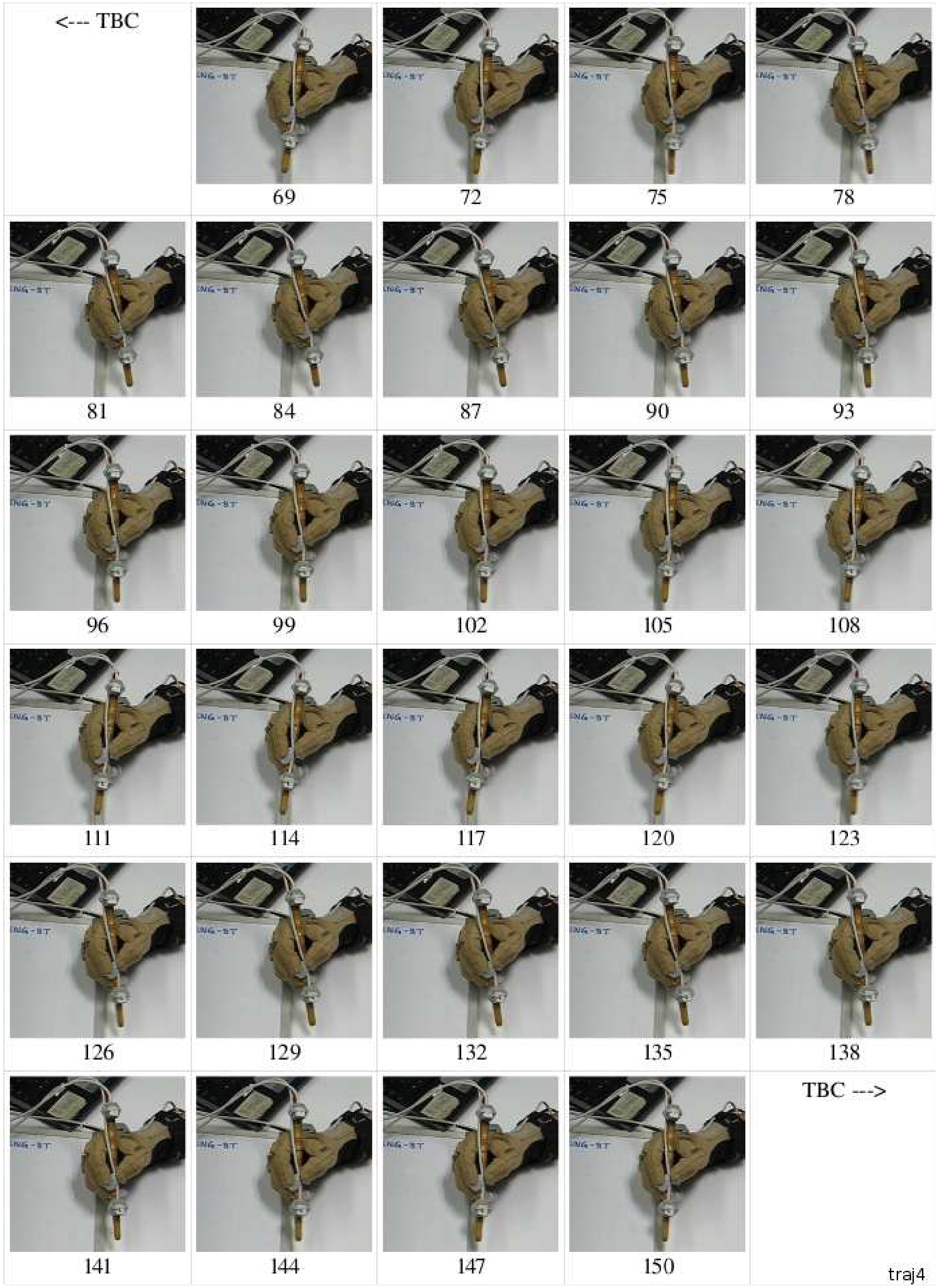


Figure 4.3: Exp. 1: Snapshot of hand movements in frame 69$^{\text{th}}$-150$^{\text{th}}$

In this experiment, dynamic tripod movement of the pen is considered. The subject were asked to perform the movement eight time consecutively. Figure 4.3 shows some portions of the movement.

3D-eigengrasp space is created using PCA on the 18 joint angles trajectory data. Figure 4.4 shows that more than 90% of variance of original data can be accounted by the first three eigengrasps.

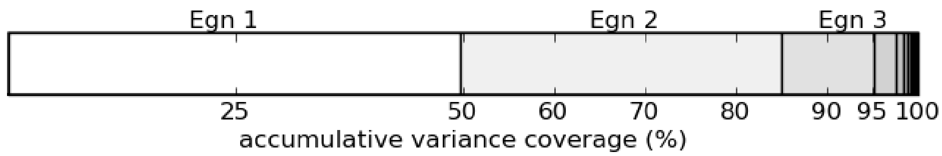


Figure 4.4: Exp. 1: Accumulative variance accounted by each eigengrasps

The 18 joint angles trajectory is projected into 3D-eigengrasp space. After some post-processing, likes smoothing, it is plotted and shown in figure 4.5 together with its curvature. Note that red dots in the trajectory are frames where its curvature are high, as shown in the corresponding dash line in figure 4.5(c).

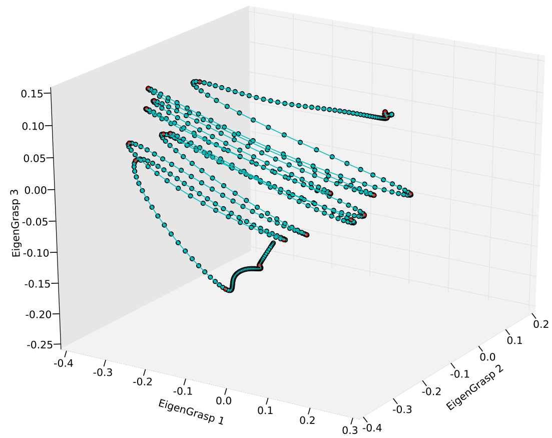
It could be noticed that in some viewpoints, we can see the linear relation of sub-trajectory which connecting together as a whole trajectory. The curvature shown in the graph are calculate from scale $s = 1, ..., 10$. It can be noticed that there are a few points marked as noises in the beginning and the end. The reason is that at those periods, hand was staying still when we were preparing to capture the data. Therefore, when the hand postures are plotted into 3D-eigengrasp space, they become too close to each others and cause some unreasonable results.

Example of some extracted intermediate grasp states are shown in figure 4.6, together with their references both in 3D-eigengrasp space and curvature graph. Some descriptions of the corresponding movement of those grasp states are described below.

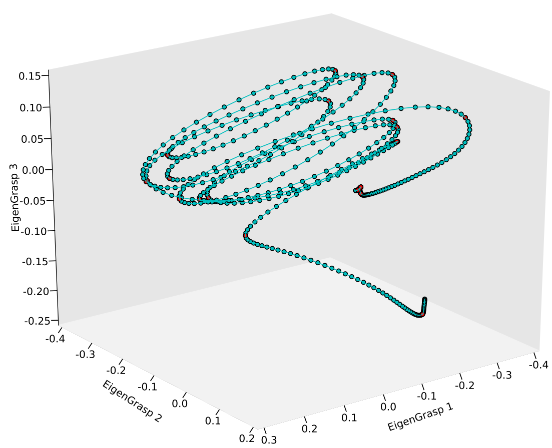
Frame 74$^{\text{th}}$ At this frame, the hand starts to bend toward the palm, causing the writing end of the pen to move down.

Frame 95$^{\text{th}}$ The hand stops moving in the same pattern described in frame 74$^{\text{th}}$, and starts to move in the opposite direction. This means that most joint angles at the finger tend to move away from the palm, causing the writing-end of the pen to move up.

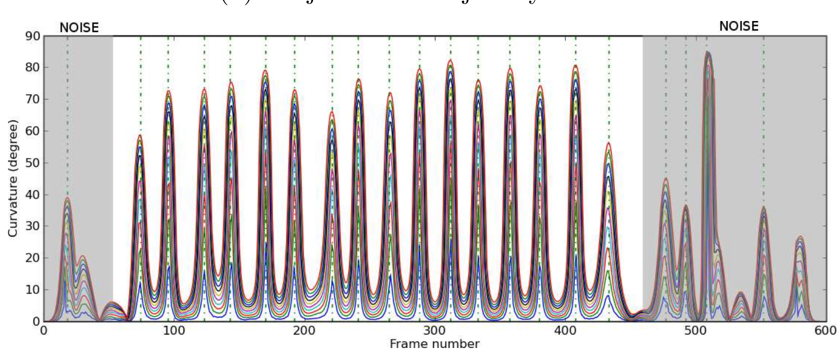
Frame123$^{\text{rd}}$ The hand stops moving in the same pattern described in frame 95$^{\text{th}}$, and starts to move in the same direction described in frame 74$^{\text{th}}$. Then, the movement keeps on going like this another seven times.



(a) Projected 3D trajectory view 1



(b) Projected 3D trajectory view 2



(c) Curvature of 3D trajectory in many scales

Figure 4.5: Exp. 1: Projected trajectory and its curvature

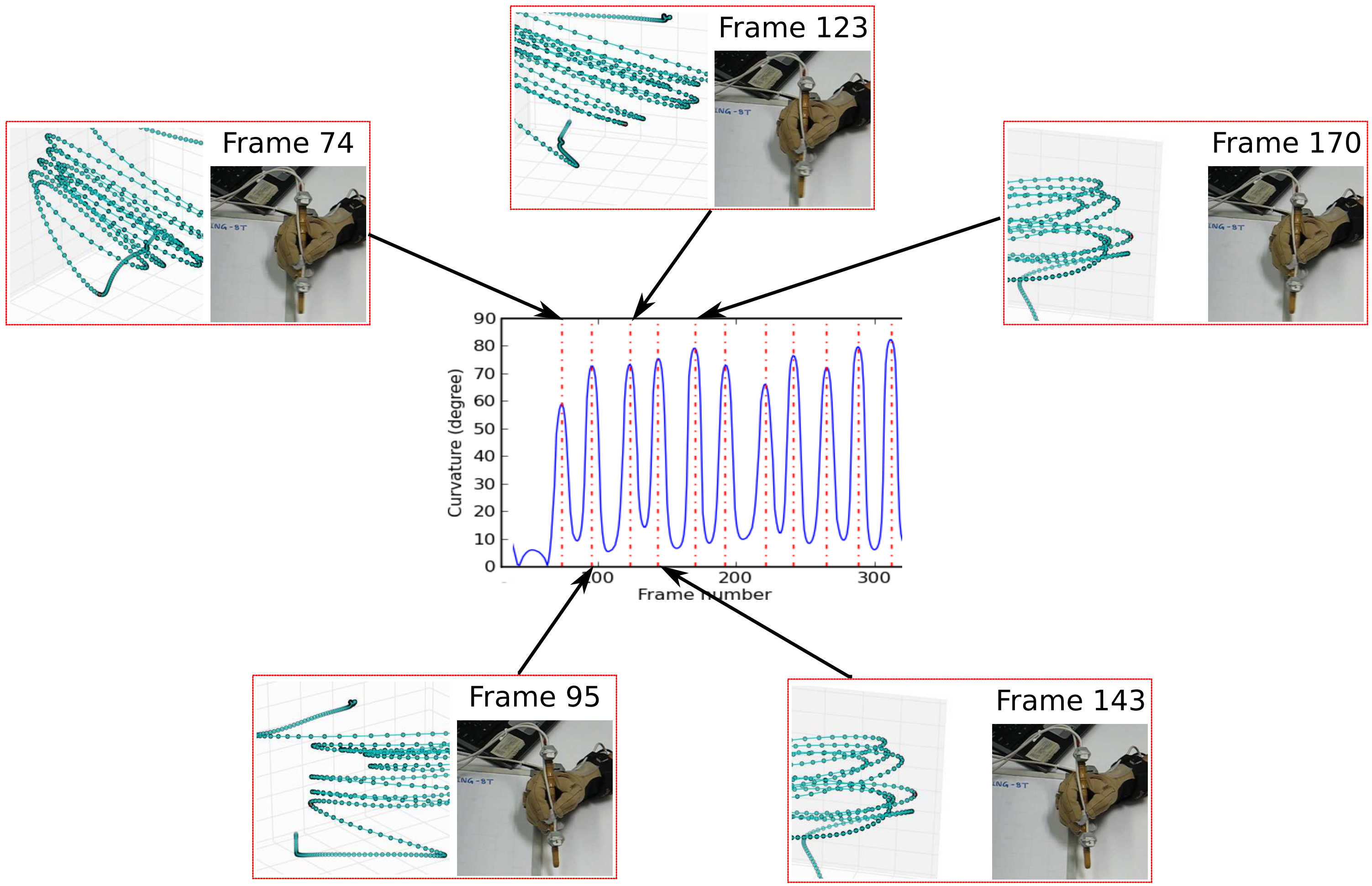
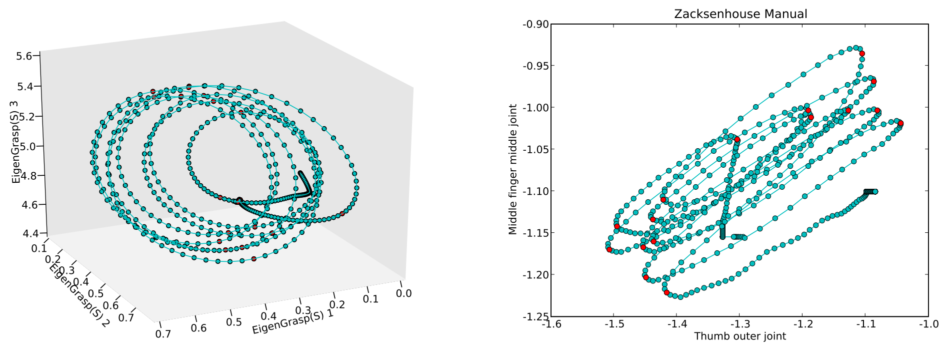


Figure 4.6: Exp. 1: Example of some extracted intermediate grasp states

Comparison

To compare the results among different experimental setups, their curvatures are shown in figure 4.8. It can be seen that most experimental setups give similar results. We analyse and compare each of them separately with experimental setup 1.

Compare with Exp. Setup 2: In Exp. Setup 2, 3D-eigengrasp space is created from hand postures captured in all data sets. The results have not shown much different in the 3D projected trajectory, curvature, and their extracted intermediate grasp states.



(a) Projected trajectory in Exp. Setup 3 (b) Projected trajectory in Exp. Setup 4

Figure 4.7: Exp. 1: Projected trajectory in other experimental setups

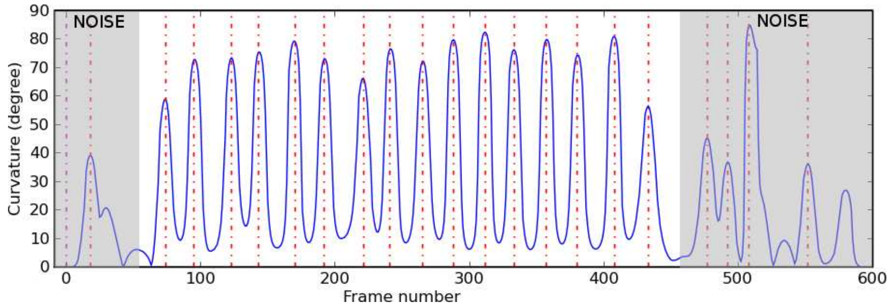
Compare with Exp. Setup 3: In Exp. Setup 3, 36 dimension vector data are used as an input data. Using PCA, the dimension is reduced to three dimensions. The dimensional reduced trajectory is shown in 4.7(a). When we consider the curvature of the trajectory, it does not have much different throughout the trajectory as the range is swung between 25 to 60 degree. Although the shape of curvature graph is different from Exp. Setup 1, unsurprisingly there are correspondence between them. We can notice that its local minimums correspond to the local maximums of Exp. Setup 1, which both are considered as intermediate grasp states. When mapping these local minimums back to the movement stream, they are corresponded to the brief stop of the hand. Other the other hand, the local maximums in this experiment are the moments where the *same-pattern* movements of the hand are at their highest speed.

Compare with Exp. Setup 4: In Exp. Setup 4, we choose *middle finger middle point* and *thumb outer joint* as two most active joints, and consider them as phrase-plane. The trajectory in this phrase-plane is shown in figure 4.7(b). We can see that the graph have highly linear relation in their sub-trajectory. After we calculate the curvature and find the local maximums, they seem to correspond to those points in Exp. Setup 1.

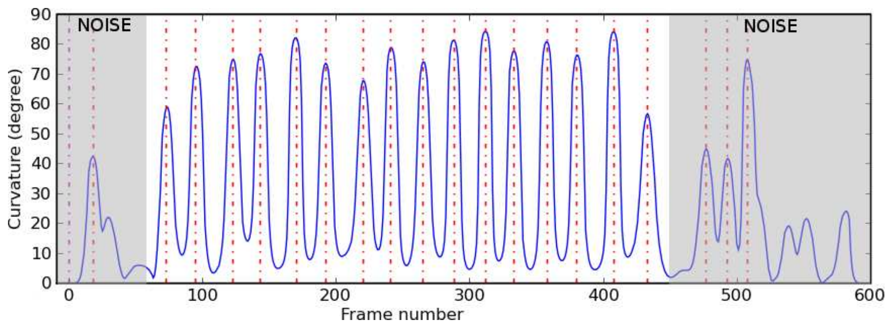
Conclusion

All experimental setups seem to work well to segment the repetition of dynamic tripod movement. Experiment setup 3 may have some problems when an automatic extraction is required, since the amplitude of the curvature graph are not very intense, even we have amplify the curvature calculation with scale $s = 10$.

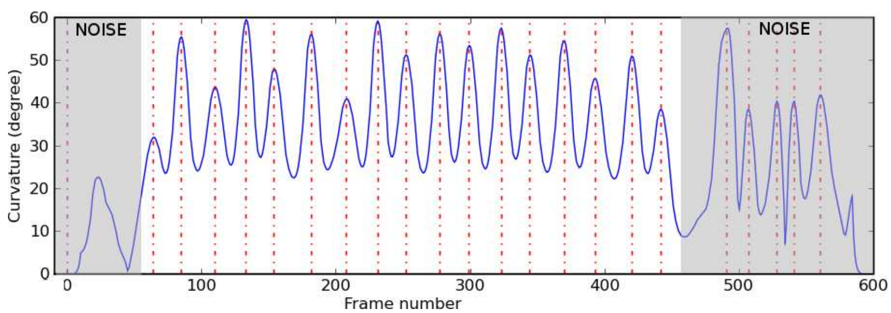
We find out that the maximum number of frames per second (110 fps) in the data acquisition system is preferred, in order to segment the movement correctly. This might be because the movement seem to be a quick movement, or in other word the period between each intermediate grasp states seem to be very short.



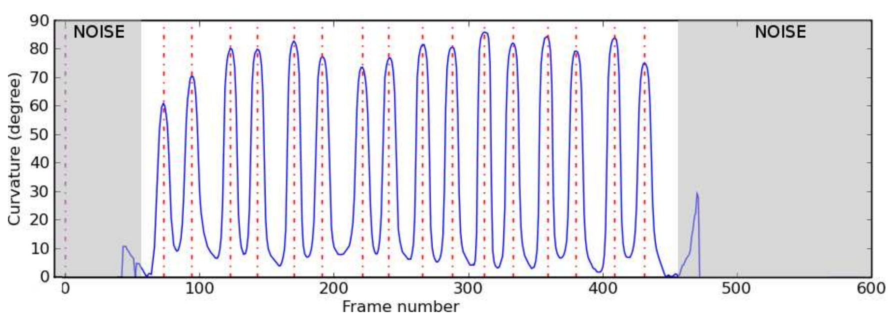
(a) Curvature from Exp. Setup 1



(b) Curvature from Exp. Setup 2



(c) Curvature from Exp. Setup 3 (local minimums)



(d) Curvature from Exp. Setup 4

Figure 4.8: Exp. 1: Comparison of curvatures with all other experimental setups

4.2.2 Experiment 2 : Rock the Rubik Cube

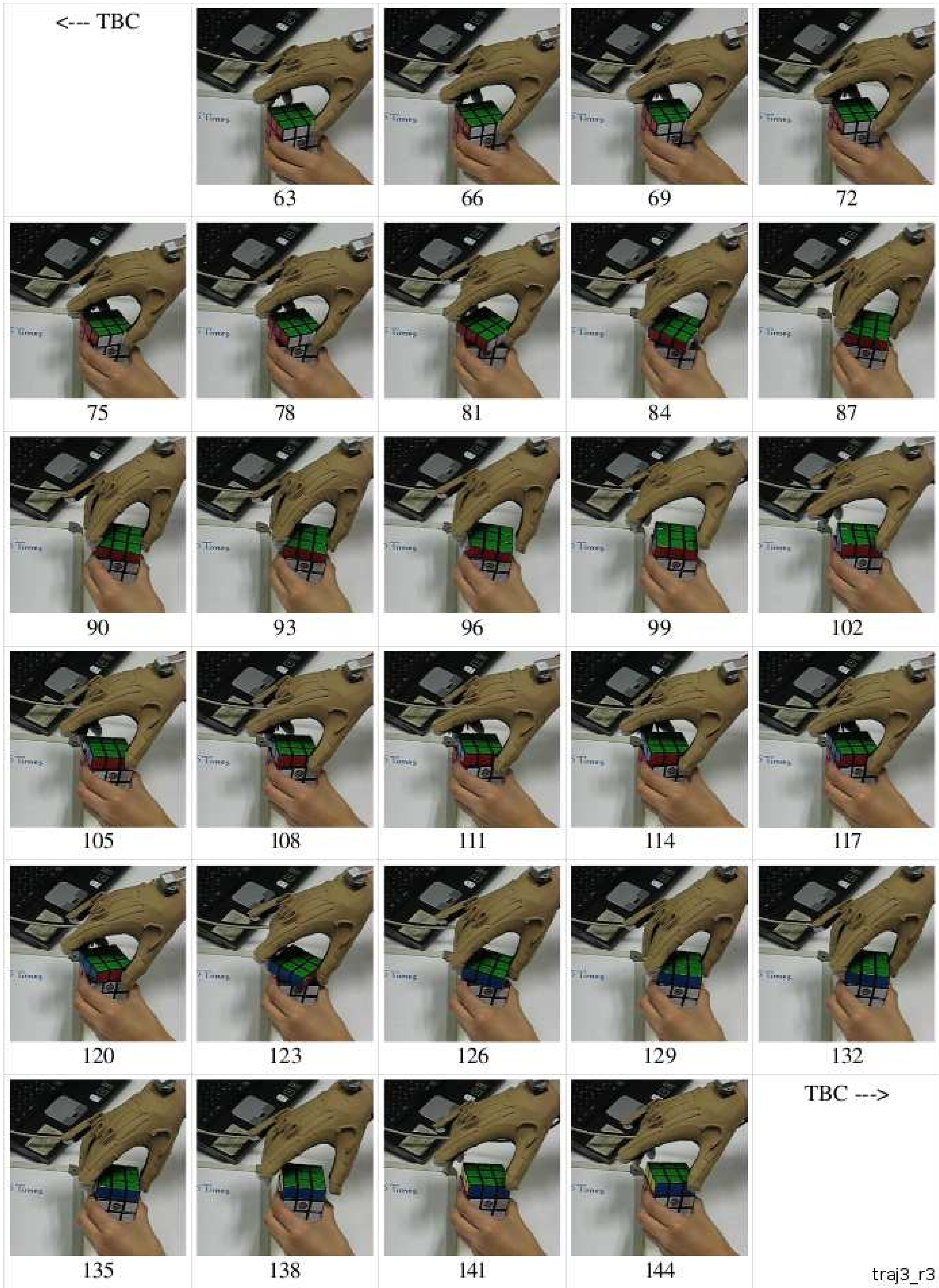


Figure 4.9: Exp. 2: Snapshot of hand movements in frame 63$^{\text{rd}}$-144$^{\text{th}}$

In this experiment rock movement is performed at Rubik cube five times consecutively. Figure 4.9 shows some portions of the movement.

PCA is used against 18 joint angles trajectory data in order to create 3D-eigengrasp space. The similar result is shown in figure 4.10; more than 90% of variance of original data is accounted by the first three eigengrasps.

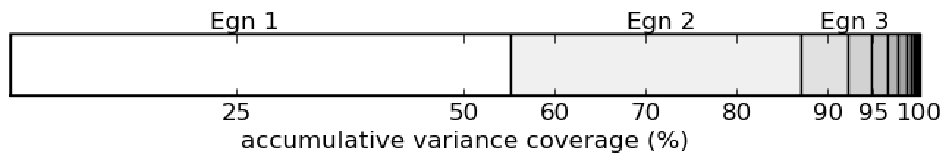


Figure 4.10: Exp. 2: Accumulative variance accounted by each eigengrasps

After some data pre-processing likes re-sampling, 18 joint angles trajectory is projected into 3D-eigengrasp space. The smoothed trajectory is shown in figure 4.11 together with its curvature. The red dots are frames where their curvature are high and are extracted as intermediate grasp states.

Similar to experiment 1, in some viewpoints the linear relation of sub-trajectory can be noticed. Curvatures shown in figure 4.11(c) are calculated from scale $s = 1, \ldots, 10$, and noisy data causing from preparation period are also segmented.

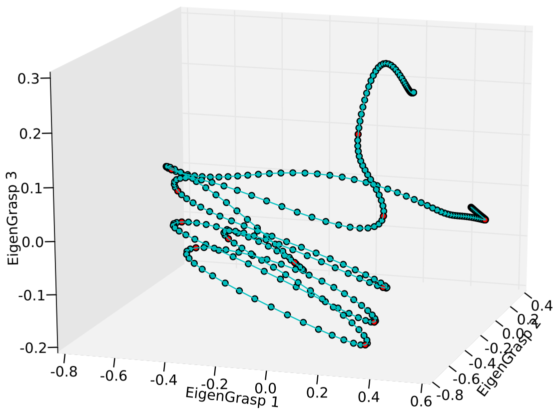
Total of 10 intermediate grasp states are extracted. Example of them are shown in figure 4.12, together with their references both in 3D-eigengrasp space and curvature graph. Some descriptions of the corresponding movement of the grasp states are describes below.

Frame 72$^{\mathbf{nd}}$ The hand grasps Rubik cube and *starts* to turn it counterclockwise. During frame 72$^{\text{nd}}$ and frame 92$^{\text{nd}}$, all joint angles tend to move in the same pattern with the purpose to turn the Rubik cube counterclockwise.

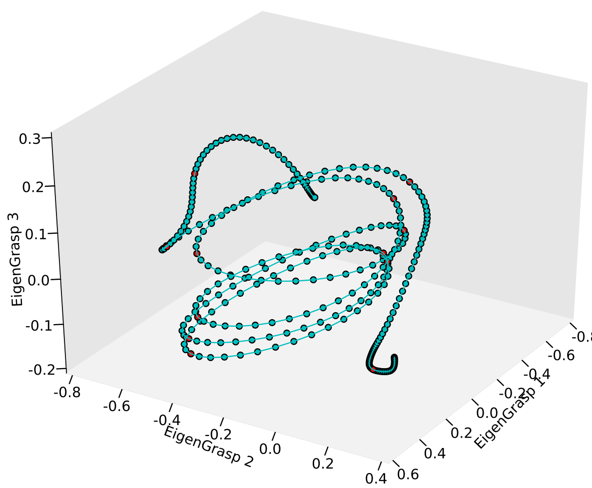
Frame 92$^{\mathbf{nd}}$ At this frame, all joint angles *stop* turning the Rubik cube, release it, and start changing the contacts as to prepare for next turn. During frame 92$^{\text{nd}}$ and frame 112$^{\text{th}}$, all joint angles move in the same pattern as to prepare a hand posture for turning a Rubik cube.

Frame112$^{\mathbf{th}}$ The hand grasps Rubik cube again and starts to repeat the action in frame 72$^{\text{nd}}$.

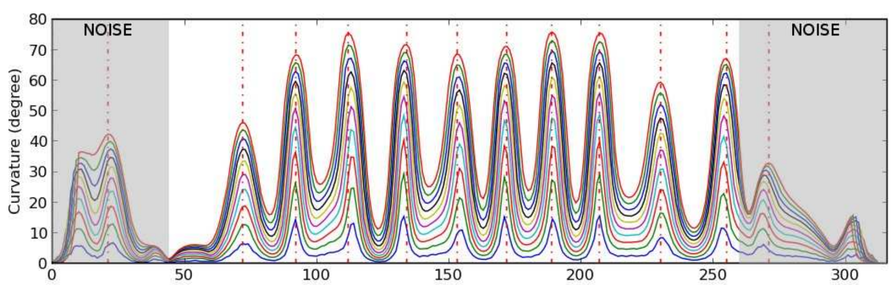
The hand continues this iteration until the Rubik cube is turned five times.



(a) Projected 3D trajectory view 1



(b) Projected 3D trajectory view 2



(c) Curvature of 3D trajectory in many scales

Figure 4.11: Exp. 2: Projected trajectory and its curvature

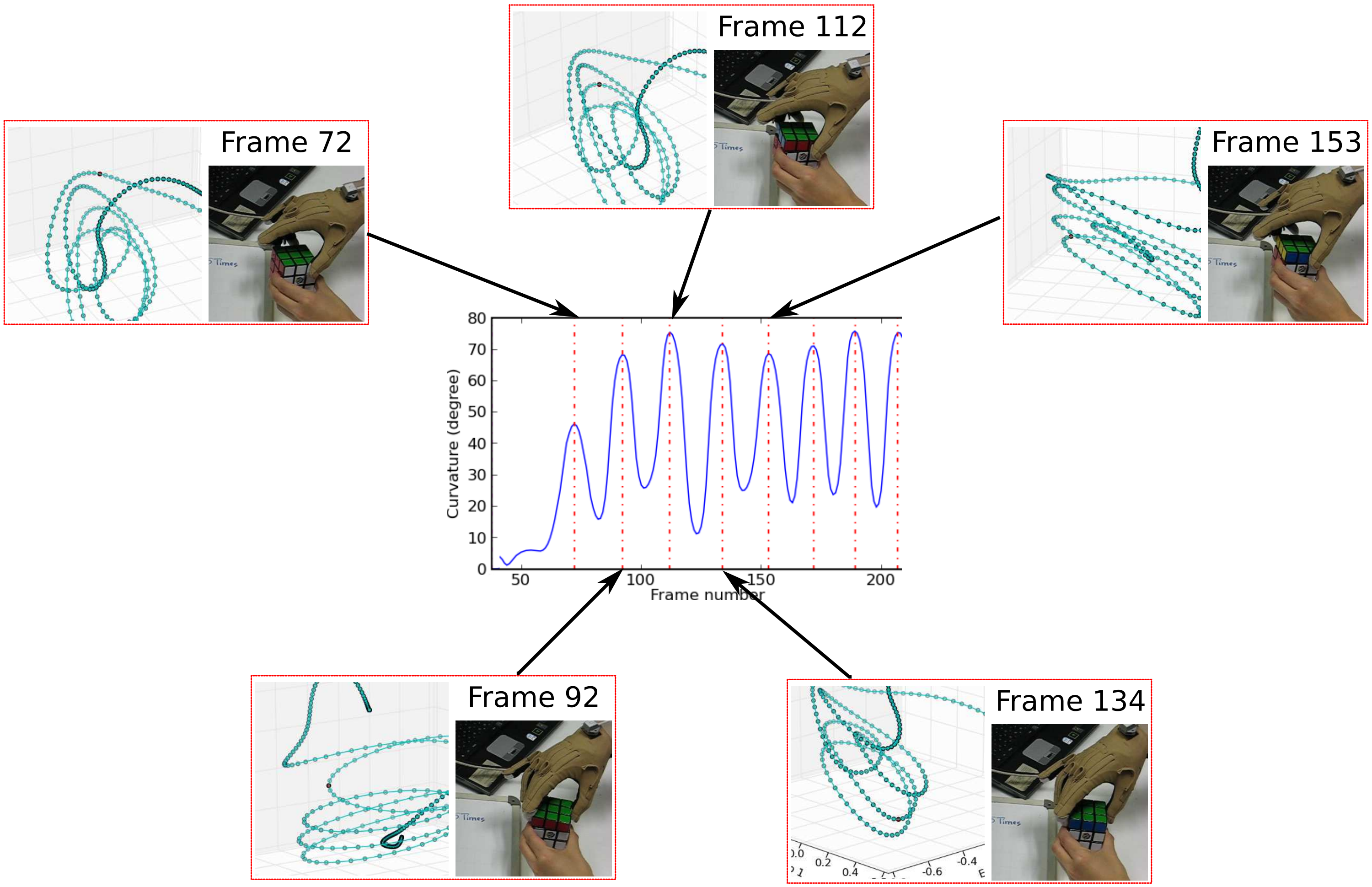
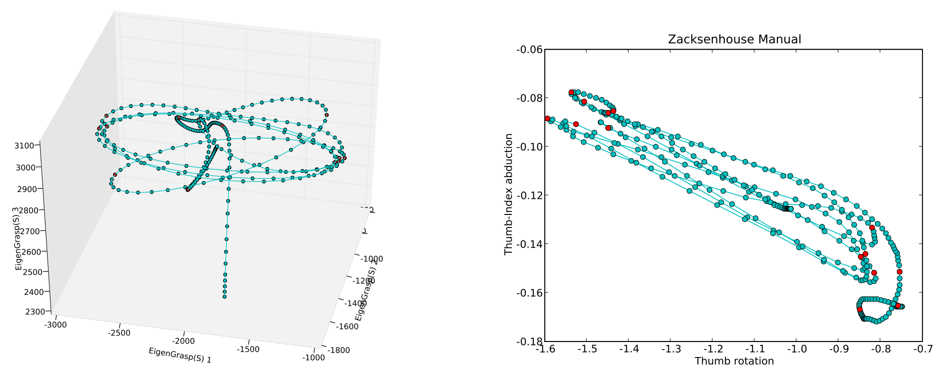


Figure 4.12: Exp. 2: Example of some extracted intermediate grasp states

Comparison

To compare the results among different experimental setups, their curvature are shown in figure 4.14. Extracted intermediate graph states from each experimental setups similar. We analyse and compare each of them with experimental setup 1.

Compare with Exp. Setup 2: As expected, the results have not shown much different between using 3D-eigengrasp space created from its own trajectory and the average of all captured data sets. 3D projected trajectory looks similar, which results in the similarity of their curvatures and their extracted intermediate grasp states.



(a) Projected trajectory in Exp. Setup 3 (b) Projected trajectory in Exp. Setup 4

Figure 4.13: Exp. 2: Projected trajectory in other experimental setups

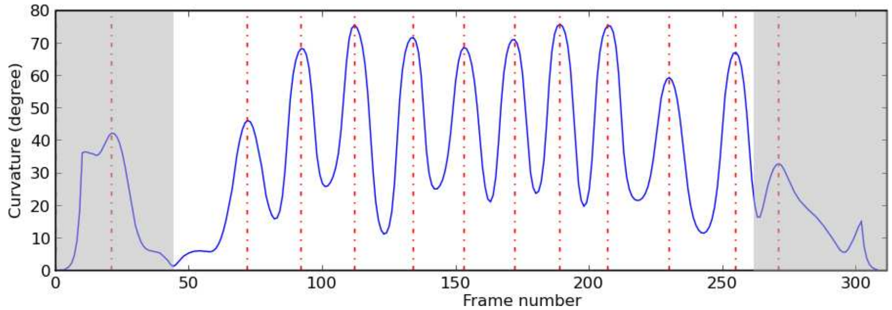
Compare with Exp. Setup 3: In Exp. Setup 3, 36 dimension vector data are used and reduced to three dimension vector. The dimensional reduced trajectory is shown in figure 4.14(b). Compared to the one in experiment 1 (dynamic tripod movement), the trajectory shows more linear relation in its sub-trajectory. In curvature graph, its local minimums also correspond to those local maximums of Exp. Setup 1. This is because the intermediate grasp states in this movement are also a brief stop of the hand.

Please not that in this experimental setup, we magnify the speed of all joint angles by 1000. This is because the speed is too small that without doing this, the graph in this experimental setup will be exactly same as the one from experimental setup 1. We do not have particular rules for choosing this, but we notice from the mean of all hand postures.

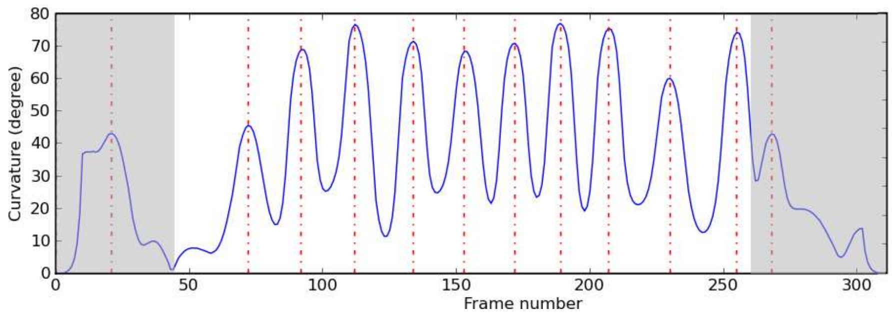
Compare with Exp. Setup 4: In Exp. Setup 4, we choose *thumb rotation* and *thumb-index abduction* as two most active joints. The trajectory in this phrase-plane is shown in figure 4.13(b), which also have highly linear relation in each sub-trajectory. In average, an extracted intermediate grasp states are similar to those extracted from Exp. Setup 1.

Conclusion

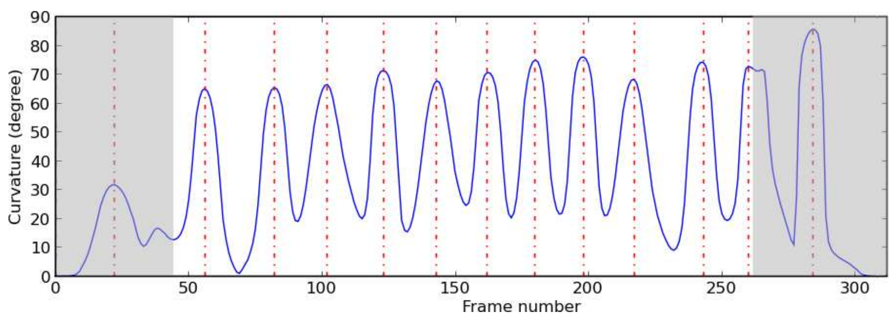
Similar to dynamic tripod movement, all experimental setups work well to segment the repetition of rock movement. A re-sampling of the captured trajectory is required as to reduce the frame per second of data acquisition system. Since the movement is done in slow manner, too high data rate may effect the extracted results as they create more noise in the calculation.



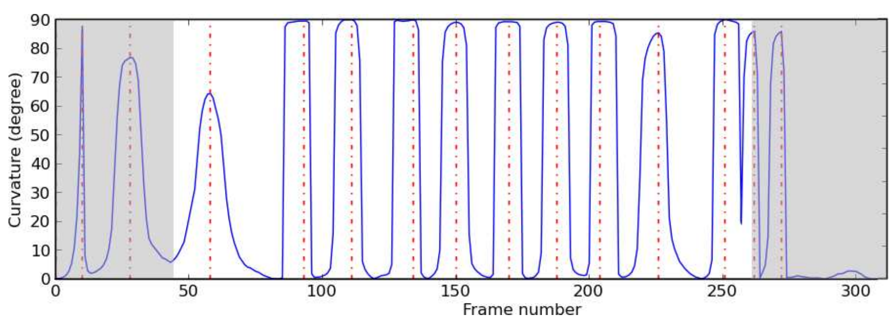
(a) Curvature from Exp. Setup 1



(b) Curvature from Exp. Setup 2



(c) Curvature from Exp. Setup 3 (local minimums)



(d) Curvature from Exp. Setup 4

Figure 4.14: Exp. 2: Comparison of curvatures with all other experimental setups

4.2.3 Experiment 3 : Interdigital Step of the Pen

Figure 4.15: Exp. 3: Snapshot of hand movements in frame 16^{th}-124^{th}

In this experiment, interdigital step movement of the pen is considered. It is considered as one of the sequential manipulative movement classified in Elliott [7], which is more complicate than simultaneous movements. Only one cycle of this movement is captured and shown in figure 4.15.

Although this trajectory is not a repetition of some movements, there is a good sign after applied PCA on 18 joint angle trajectory data to create 3D-eigengrasp. Figures 4.16 shows that more than 90% of variance of original data can be accounted by the first three eigengrasps.

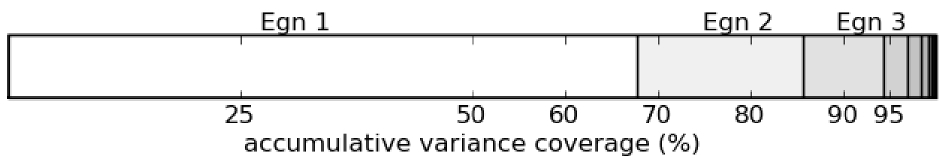


Figure 4.16: Exp. 3: Accumulative variance accounted by each eigengrasps

The 18 joint angles trajectory is down-sampled, and projected into 3D-eigengrasp space. After some smoothing process, the dimensional reduced trajectory is plotted as shown in figure 4.18, together with its curvature. The red dots are extracted intermediate grasp states, where their curvature are high in curvature graph.

The linear relation in the trajectory may not as clear as prior experiments which are a repetition of some movement, but it can be seen in some correct perspectives. The curvature are calculated from scale $s = 1$ to $s = 10$, and shown in figure 4.18(c). The intermediate grasp states are extracted as local maximums in the curvature calculated using highest scale. Four to Six of them are extracted, depending on which frames are consider to be the beginning and the end of the trajectory. Five of hand postures of an extracted grasp states are shown in figure 4.19, together with their references both in 3D-eigengrasp space and curvature graph. Their descriptions are given below.

Frame 20$^{\text{th}}$ This frame is a beginning frame where thumb, index and middle finger start to stretch out to to create room for manipulating a pen. During frame 20$^{\text{th}}$ and frame 45$^{\text{th}}$ are all devoted for this stretching.

Frame 45$^{\text{th}}$ After the space is created, thumb tries to move its tip to the under position where it can push the pen up. During the thumb movement, index and middle are stay still maintaining the pen while ring and little finger stretch out a little bit.

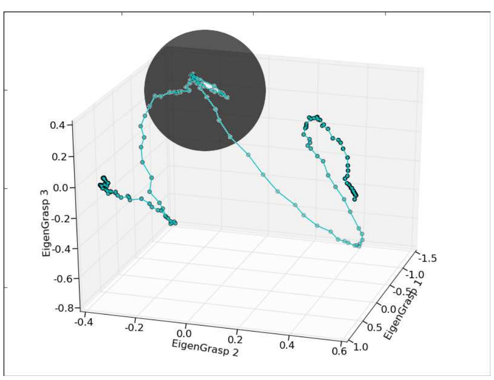
Frame 70$^{\text{th}}$ While all other fingers stay still, thumb starts to use its tip to push a pen forward. Before it stops, the thumb slides a little bit to the side of the pen preparing for next movement.

Frame 95$^{\text{th}}$ Thumb starts to use its side to push the pen against middle finger, which causes the writing end of the pen to move up. At the same time, index finger and ring finger tend to bend backward to give space to the pen which is moving.

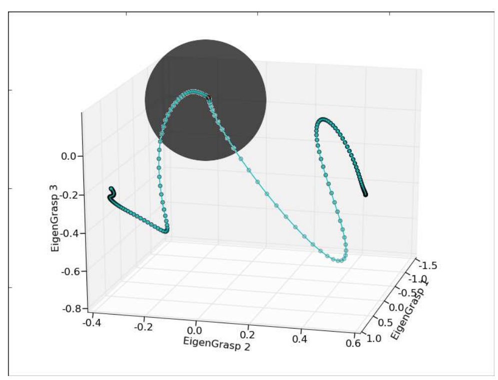
Frame 135$^{\text{th}}$ The pen are stopped when the hand fully employs the pen with drawing grasp type 2. During frame 135$^{\text{th}}$ and frame 164$^{\text{th}}$, while graping the pen tightly, the wrist rotates to the final position.

Frame 164$^{\text{th}}$ This is the frame where the hand stop its movement completely. It can be noticed that there are some frames afterward, because a delay occurred before stop capturing. This can be solved easily be better segmentation system.

Note that during frame 70$^{\text{th}}$ and frame 95$^{\text{th}}$, there should be another intermediate grasp state in frame 86$^{\text{th}}$. This is the frame where the thumb actually stops pushing a pen forward, and start to slide to the side of the pen. The proposed method miss to extract this point because of the loss of data in the smoothing process. Figure 4.17 shows the projected trajectory both before and after smoothing. We can see in the grey circle in figure 4.17(a) that there are still some more points which can be extracted as intermediate grasp states, because they connect between two linear sub-trajectory. However, these points are disappear after the trajectory is smoothed. This problem might be resolved by using better smoothing techniques.

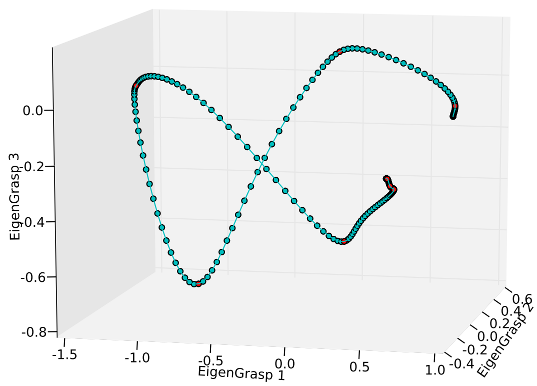


(a) Projected trajectory before smoothing

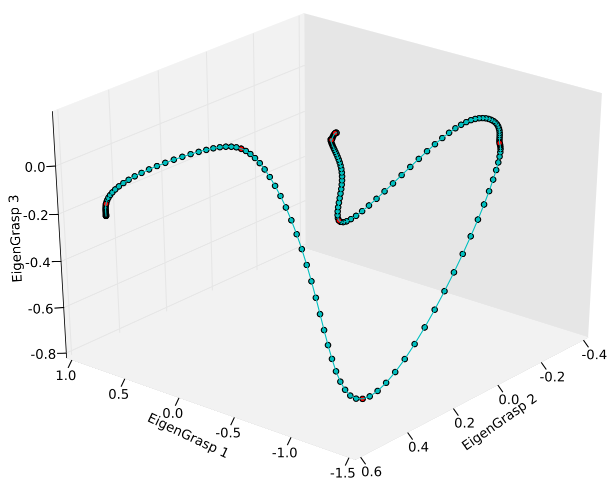


(b) Projected trajectory after smoothing

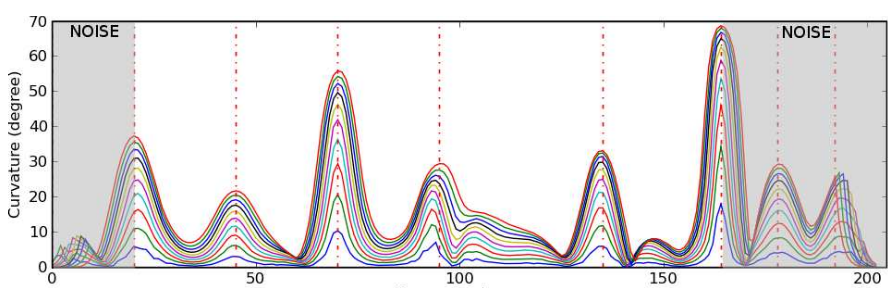
Figure 4.17: Exp. 3: Data loss due to trajectory smoothing



(a) Projected 3D trajectory view 1



(b) Projected 3D trajectory view 2



(c) Curvature of 3D trajectory in many scales

Figure 4.18: Exp. 3: Projected trajectory and its curvature

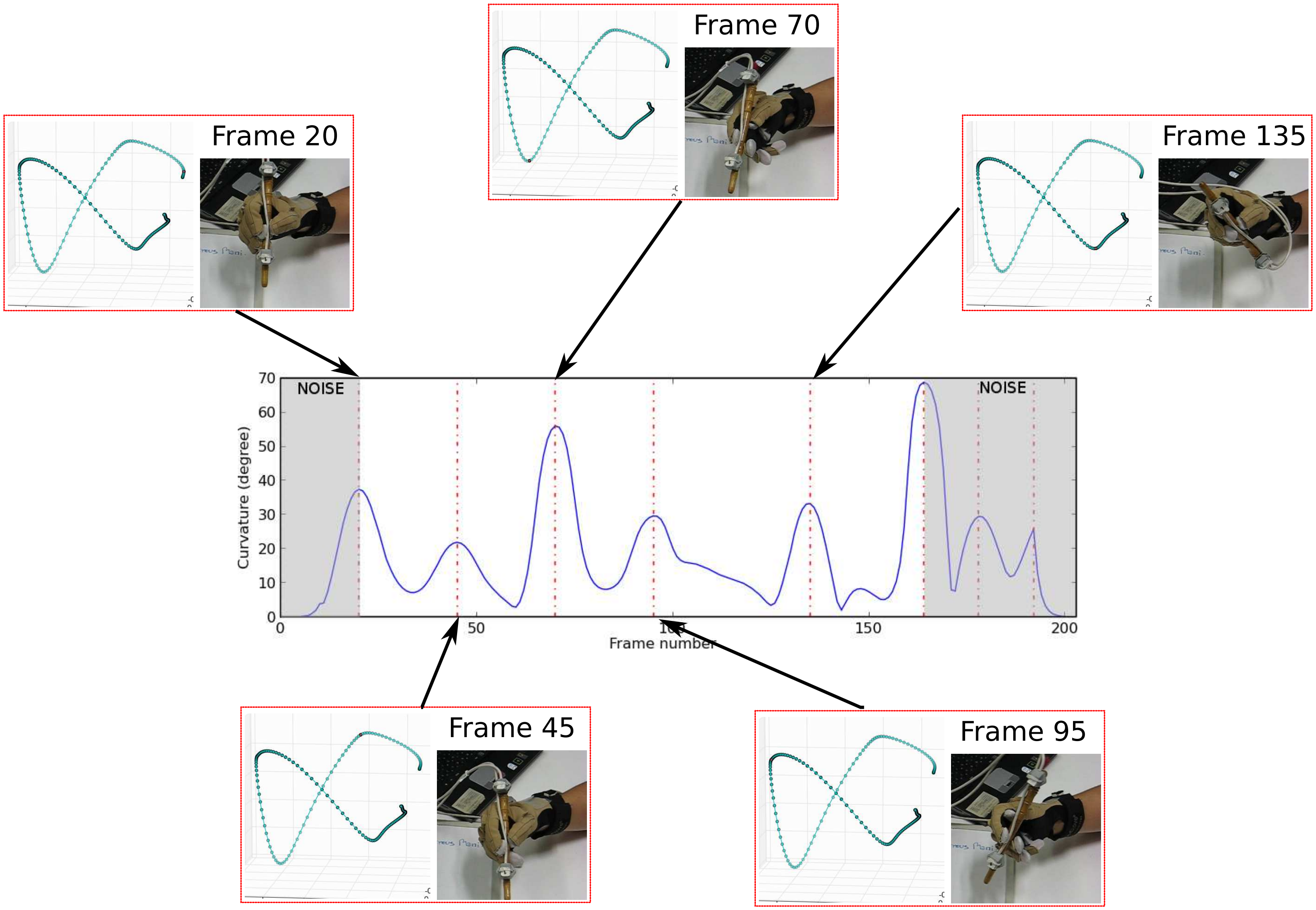


Figure 4.19: Exp. 3: All extracted intermediate grasp states

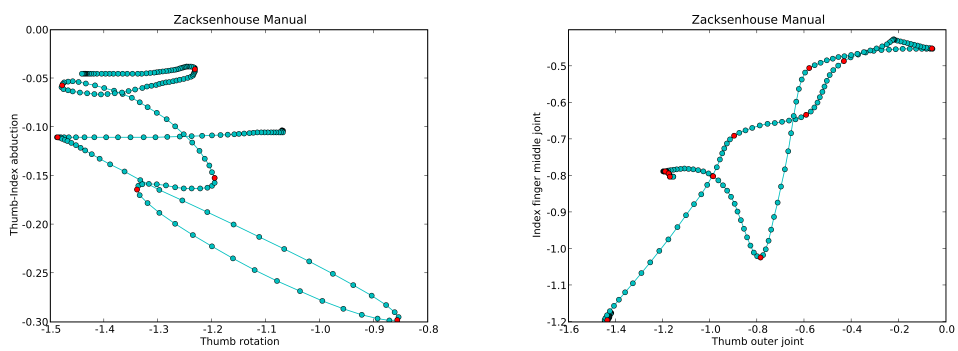
Comparison

To compare the results among different experimental setups, their curvatures are shown in figure 4.21. However, in this experiment two experimental setup 4 are conducted using different phrase-planes. Therefore, their results are shown instead of the one from experimental setup 2 which is omitted. We analyse and compare each of them separately with experimental setup 1.

Compare with Exp. Setup 2: As expected, the 3D projected trajectory, curvature ,and the extracted intermediate grasp states are very similar to those in Exp. Setup 1.

Compare with Exp. Setup 3: In Exp. Setup 3, 36 dimension vector data are used and reduced to three dimension vector using PCA. Curvature are calculated and shown in figure 4.21(b). It can be seen that there are some correspondences between the local minimums in Exp. Setup 3 and the local maximums of Exp. Setup 1, which both are the extracted grasp states in each Exp. Setups. This is because Exp. Setup 3 extracts all brief stop points in the demonstration stream. However, since the meaning of dimensional reduced trajectory is not clear, it is very difficult to compare and explain what is the different between each local minimums which have different curvature's values.

Please note that with the same reason given in experimental 2 (rock the Rubik cube), we also magnify the speed of all joint angles by 1000.



(a) Projected trajectory in Exp. Setup 4 (Thumb rotation – Thumb index abduction)

(b) Projected trajectory in Exp. Setup 4 (Thumb outer joint – Index finger middle joint)

Figure 4.20: Exp. 3: Projected trajectory in experimental setup 4, when considered two different phrase-planes

Compare with Exp. Setup 4: In Exp. Setup 4, two different phrase-plane are considered. In each phrase-plane, two most active joint angles are chosen manually. Both their trajectories are plotted and the curvatures are calculated.

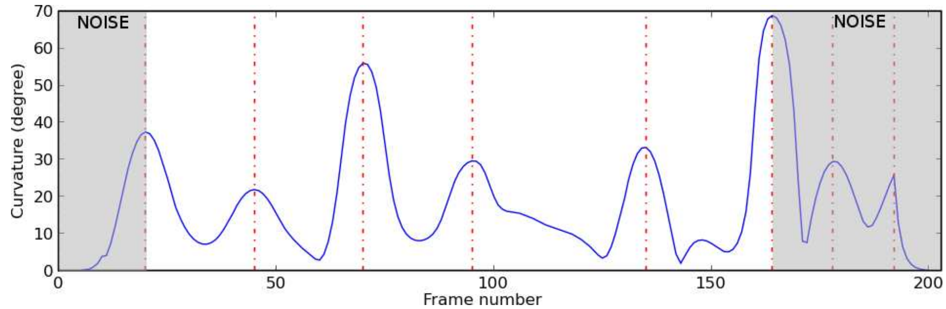
Figure 4.20(a) and 4.21(c) show the trajectory and curvature when consider *thumb rotation* and *thumb-index abduction* as most active joint angles. We can see that the trajectory in this phrase plane shows very high linear-relation in their sub-trajectory. The curvature is computed and the local maximums are extracted as intermediate grasp states. When we compare the result with the one from Exp. Setup 1, they seem to be corresponded to each others. The result from this Exp. Setup 4 even gives better result around frame 95^{th}, where Exp. Setup 1 suffered from data loss due to trajectory smoothing. However, when we look around frame 135^{th}, Exp. Setup 4 seems to give incorrectly results as they extract frame 122^{nd} and 155^{th} instead. This result is understandable, since during these period the joint angles that are active are not those of the thumb which considered as phrase-plane, but they are joint angles of index finger and the wrist.

To verify our assumption above, another phrase-plane is considered. In the second experiment of Exp. Setup 4, we choose *thumb outer joint* and *index finger middle joint* as the most active joint angles. The result is shown in figure 4.20(b) and 4.21(d). We can see that this phrase-plane give correct result at frame 135^{th} where the index finger is active, but it give incorrect results elsewhere. This is because index finger actually move at those frames, but it is not the main joint angles that play an important role in manipulating the pen.

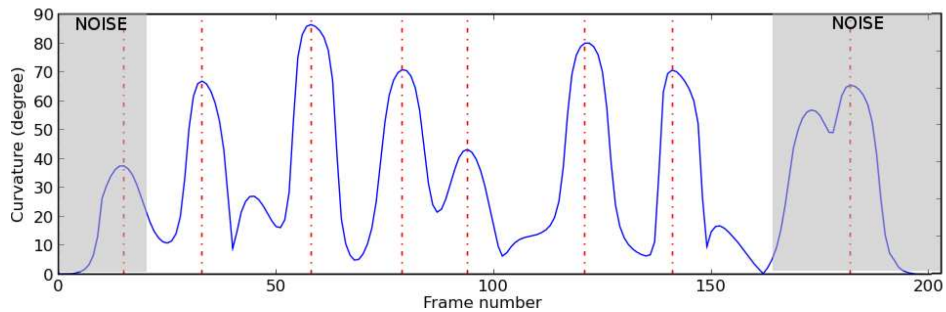
Conclusion

In this experiment, we consider more complex movement. There are some differences among each experimental setups.

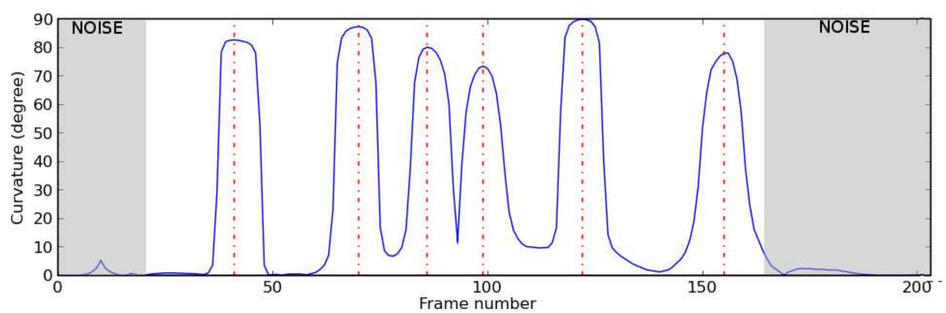
Exp. Setup 1 and Exp. Setup 2 seem to suffer from information loss during the smoothing processing of 3D trajectory. While Exp. Setup 3 provides similar results with Exp. Setup 1, the explanation of both the trajectory and curvature are still unclear, and a further analysis of the meaning of the principle components is necessary. In Exp. Setup 4, it shows a reasonable result when the phrase-plane is chosen correctly. However, more than one phrase-plane may be necessary in the complex movement when there are many active joint angles in the manipulation.



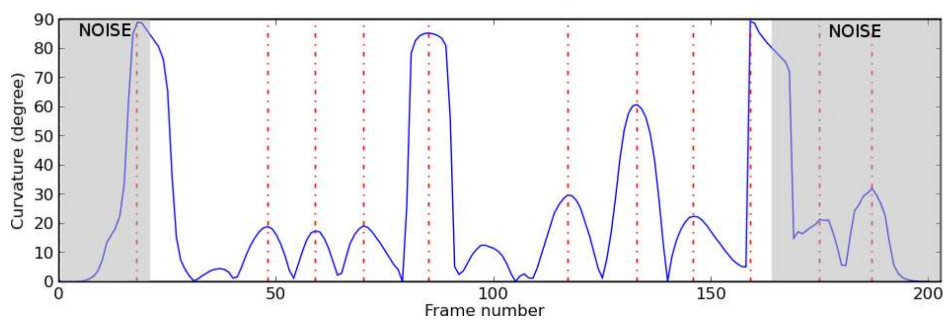
(a) Curvature from Exp. Setup 1



(b) Curvature from Exp. Setup 3 (local minimums)



(c) Curvature from Exp. Setup 4 (Thumb rotation – Thumb index abduction)



(d) Curvature from Exp. Setup 4 (Thumb outer joint – Index finger middle joint)

Figure 4.21: Exp. 3: Comparison of curvatures with other experimental setups

4.3 Discussion

Three different movements of dexterous manipulation have been considered in order to to verify our proposed method. We have conducted four experiment setups, including the proposed method, to order to compare the results. We find out that in the repetition of simple movement, like those in section 4.2.1 and 4.2.2, the result from all experiment setups are very similar. However, it is not the case when the manipulative movement becomes more complex, like the one in section 4.2.3.

Furthermore, when the movement are more complex and not the repetition of simple movement, we have to admit that we still lack of more systematic measurement for evaluation. Although we have tried many methods to extract intermediate grasp states in order to compare the results, the final judgement that decides which results are good or bad is still based on the experimenter himself/herself.

Based on our segmentation's criteria given in section 1.3.2 and the fact that the only evaluation we have now for the complex movement is the experimenter's judgement, we describe some characteristics that would effect the results of each experimental setups.

Exp. Setup 1 (the proposed method) and Exp. Setup 2 seem to have similar result in all experiments. This mean that it does not matter whether to use its own trajectory or the average trajectory of same movements to create 3D-eigengrasp space. The results of both Exp. Setup 1 and 2 are also affected by very small changes in the movement. The problem for this is that when the trajectory is projected into 3D-eigengrasp, these small changes will be represented by a short linear sub-trajectory, and then when the smoothing techniques is applied on them, they disappear. This problem might be resolved by a better smoothing techniques.

Exp. Setup 3 also shows good results in all experiments, especially in experiment 1 and 2. The downside of this method is that when the dimension of input data are reduced using PCA, the meaning of each principle components becomes ambiguous, and it becomes more difficult when the analysis of the extracted grasp states is needed. Furthermore, since the assumption of this method is to detect the brief stops in the stream of demonstration, it is still doubtful about its performance when the movement are pretty fast, like the one in experiment 1. Therefore, in order to extract intermediate grasp states in fast movement, faster data acquisition system might be necessary.

The method used in Exp. Setup 4 is actually implemented based on Zacksenhouse [19]. It also has good results in all experiments. However, the criteria for choosing two most active joints is necessary, since the phrase-plane that performs well in each manipulative movements are still different.

Moreover, when the movement become more complicated and each groups of joint angles effect in different small movement like the one in experiment 3 (interdigital step), an analysis of more than one combination of two active joints may be necessary. However, by doing so it may bring up another difficult problem which is to choose the intermediate grasp states among different phrase-planes.

In all Exp. Setups, the results are still highly affected by some parameters. For example, a number of appropriate frames per second required for each particular methods and movements, or smoothing parameters used in Exp. Setup 1 which effect a detection of small movement etc. This means that a good approach for choosing these parameters is needed in order to automatize both the proposed method and others.

From all above, rather than using a subjective evaluation, a more concrete method to evaluate the result is necessary. Otherwise, we will not be able to tell which results are good, and which is not.

CHAPTER 5

Conclusion

This thesis describes a preliminary step towards automatic dexterous manipulation of robot from observation. Firstly, an object manipulation in the context of programming by demonstration is briefly explained. Then dexterous manipulation is introduced into the existing system[5]. We suggest that a first step for dexterous manipulation planning in programming by demonstration is to segment a stream of human manipulative movement into shorter meaningful sequences.

Normally, segmenting meaningful sequences in dexterous manipulation would refer to defining the moments in the manipulative movement where contact relation between hand and grasping object is changing. However, there is a limitation to detect and recognize those moments due to the current technology of tactile sensors. Therefore, a method to detect the moments where contact relation is changing by considering a movement of the hand is proposed. This is considered to be the main contribution of this thesis.

A segmentation is done based on the assumption that contact relation between hand an grasping object would change when a coordinative movement of the hand changes. By saying a coordinative movement of the hand, we refer to a period when all joint angles of the hand are moving coordinately in some particular pattern without changing throughout the movement. A segmentation is done in the reduced dimensional joint space of the hand. Principle component analysis is used to reduce a highly dimensional joint space into three dimensional space. In this reduced space, a trajectory of coordinative movements of the hand is turn out to be approximately linear. A local maximums in curvature of the dimensional-reduced joint trajectory is points where trajectory dramatically changes its direction. These points can equally refer as the moments where there are a change in the pattern of a coordinative movement of the hand. Therefore, curvature property of the joint trajectory in reduced three dimensional joint space is used as a segmentation criteria.

As for an evaluation, three dexterous manipulative movements are considered as experimental data. Two of which are the repetition of the short movement which referred as simultaneous hand movement, while the other is more complex movement that is usually referred as a type of sequential manipulative movement. Each experimental data is processed with the proposed method and another three different segmentation methods. All results are compared and discussed.

5.1 Future Works

Future works can be divided into two separate groups; an improvement of current method and a future direction of research.

As for an improvement of current proposed method, the future work is already discussed and proposed in 4.3 that a better measurement to evaluate the results is necessary, especially for a complex manipulative movement. This should be considered as a first priority to improve the efficiency of the proposed method, because without the measurement, it is impossible to judge whether the segmentation results is correct or not.

As for a future direction of research, there are two directions we would like to proposed.

Research Direction 1

> Based on the belief that human has particular type of grasp for specific type of objects [8], we may be able to further extend this concept with human movement as well. *For a specific type of objects or category of objects, human may have particular manipulative movements for them.* If this is true, we may be able to use these movements to create a global 3D-eigengrasp space for specific category of objects.
>
> What is special about this is that, the performance of the proposed method may increased by doing this. Referring back to section 4.2.3, there is some movements that is cancelled out in the smoothing processing after the trajectory is projected into 3D-eigengrasp space. The smoothing techniques may not be the only reason for it. Another reason might be because the cancelled out movement may only create a tiny change when it is projected into 3D-eigengrasp space.

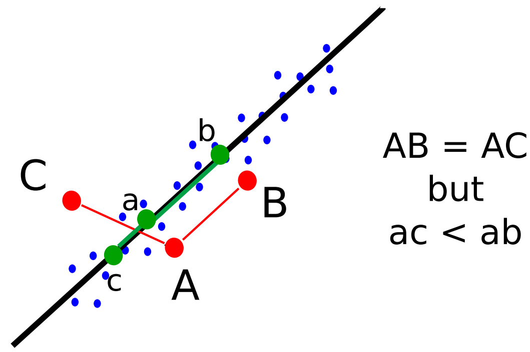


Figure 5.1: A difference in length after projection

Figure 5.1 shows that lines with equally length can be different after projection. It can be seen that $\overline{AB}$ is in the direction of projection axis, so that after projection the length of its projection, $\overline{ab}$, does not change when compare to $\overline{AB}$. However, on the other hand the length of line $\overline{ac}$ is considerably short when compared to $\overline{AC}$. This is because direction of $\overline{AC}$ is not along the projection axis.

The same understanding can be applied in the joint space as well. If the hand postures of the movement has affected in the creation of 3D-eigengrasp space, when a movement is projected into 3D-eigengrasp space, the initial and final hand postures of the movement may result in a longer trajectory.

Once after the above condition is archived, we may change to different type of objects. As it might notices that in our experiment, we considered greatly about pen-like objects. We believe that our proposed method should be able to applied in any kind of objects. Therefore, as for this research direction a variation of objects should also be further explored.

Research Direction 2

An automation of dexterous manipulation from observation is also another interesting topic. To enable a robot to solely observe human demonstration and perform a dexterous manipulation on it own with least human intervention , the following steps might be necessary.

1. A method to choose appropriate parameters in our proposed method. It might be thought that our proposed method for intermediate grasp states extraction can be used without human intervention.

However, there are a lot parameters that is still manually setted by the operators, for example parameters for smoothing process, number of suitable frames per second, threshold to decides a local maximums of curvature curve etc.

2. After the intermediate grasp states have been discover, a more appropriate representation than a joint angles of the hand postures is necessary. The most appropriate representation for this would be the contact relation between a hand and the grasping object. An approach for this is to utilize all information getting from a current sensors (hand posture, hand and object position) and map them into a pre-defined model, and then estimate the contact relation out of it. The pre-defined model can also be a simulation, which would allow more of the visualization. However, if this is still not enough to decide where exactly the contact points are, one might consider adding a position pre-defined tactile sensors into data acquisition system.

3. Finally, before mapping the contact point representation we extracted from human demonstration to the real robot hand, we may have to consider physical constraints. This is because it cannot be guaranteed that a degree of freedom of the target hand would be able to perform exactly as the representation tell it to do. As for this problem, the contact point representation may have to be revised in order to connect the physical world with the planning, or another layer of representation, like virtual finger [16], may have to be considered.

BIBLIOGRAPHY

[1] Jun Takamatsu, Koichi Ogawara, Hiroshi Kimura, and Katsushi Ikeuchi, "Recognizing Assembly Tasks Through Human Demonstration," *International Journal of Robotics Research*, vol. 26, no. 7, pp. 641–659, 2007.

[2] J. R. Napier, "The prehensile movements of the human hand," *The Journal of Bone and Joint Surgery*, vol. 38-B, no. 4, pp. 902–913, Nov 1956.

[3] M.R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 269–279, Jun 1989.

[4] T. Iberall, "The nature of human prehension: Three dexterous hands in one," *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, vol. 4, pp. 396–401, Mar 1987.

[5] Sing Bing Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-temporal segmentation of tasks from human hand motion," *Robotics and Automation, IEEE Transactions on*, vol. 11, no. 5, pp. 670–681, Oct 1995.

[6] S. Kudoh, N. Ikeda, K. Ogawara, and K. Ikeuchi, "Learning Everyday Object Manipulation from Observation," *Proc. IROS'08 Workshop on Grasp and Task Learning by Imitation*, September 2008.

[7] J.M. Elliott and Connolly K.J., "A classification of manipulative hand movements," *Developmental medicine and child neurology*, vol. 26, no. 3, pp. 283–296, Jun 1984.

[8] N. Kamakura, M. Matsuo, H. Isshi, F. Mitsuboshi, and Y. Miura, "Patterns of static prehension in normal hands," *The American journal of occupational therapy*, vol. 34, no. 7, pp. 437–445, Jul 1980.

[9] C.C. Kemp, A. Edsinger, and E. Torres-Jara, "Challenges for robot manipulation in human environments; Grand Challenges of Robotics," *Robotics & Automation Magazine, IEEE*, vol. 14, no. 1, pp. 20–29, March 2007.

[10] K. Ikeuchi and T. Suehiro, "Toward an assembly plan from observation. I. Task recognition with polyhedral objects," *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 3, pp. 368–385, Jun 1994.

[11] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, and K. Ikeuchi, "Task model of lower body motion for a biped humanoid robot to imitate human dances," *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pp. 3157–3162, Aug. 2005.

[12] H. Liu, T. Iberall, and G.A. Bekey, "The multi-dimensional quality of task requirements for dexterous robot hand control," *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on*, pp. 452–457 vol.1, May 1989.

[13] Sing Bing Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-recognizing a grasp from observation," *Robotics and Automation, IEEE Transactions on*, vol. 9, no. 4, pp. 432–443, Aug 1993.

[14] A.M. Okamura, N. Smaby, and M.R. Cutkosky, "An overview of dexterous manipulation," *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 1, pp. 255–262 vol.1, 2000.

[15] K. Bernardin, K. Ogawara, K. Ikeuchi, and R. Dillmann, "A sensor fusion approach for recognizing continuous human grasping sequences using hidden Markov models," *Robotics, IEEE Transactions on*, vol. 21, no. 1, pp. 47–57, Feb. 2005.

[16] M.A. Arbib, T. Iberall, and D.M. Lyons, "Coordinated control programs for movements for the hand," in *Hand Function and The Neocortex*, W. Goodwin and Darian-Smith, Eds., pp. 111–129. Springer-Verlag, Berlin, 1985.

[17] Marco Santello, Martha Flanders, and John F. Soechting, "Postural Hand Synergies for Tool Use," *The Journal of Neuroscience*, vol. 18, no. 23, pp. 10105–10115, December 1998.

[18] T. Iberall, G. Bingham, and M.A. Arbib, "Opposition space as a structuring concept for the analysis of skilled hand movements," in *Generation and Modulation of Action Patterns*, H. Heuer and C. Fromm, Eds., pp. 158–173. Springer-Verlag, Berlin, 1986.

[19] M. Zacksenhouse and T. Moestl, "Segmenting manipulative hand movements by dividing phase plane trajectories," *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, vol. 3, pp. 1910–1915 vol.3, 1999.

[20] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dimensionality reduction for hand-independent dexterous robotic grasping," *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pp. 3270–3275, 29 2007-Nov. 2 2007.

[21] I. S. Lindsay, "A tutorial on Principle Components Analysis," www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf, February 2002.

[22] I. Dejmal and M. Zacksenhouse, "Coordinative Structure of Manipulative Hand-Movements Facilitates Their Recognition," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 12, pp. 2455–2463, Dec. 2006.