Electronic Thesis and Dissertation Repository

11-15-2013 12:00 AM

A Hybrid Visual Control Scheme to Assist the Visually Impaired with Guided Reaching Tasks

Duane Jacques
The University of Western Ontario

Supervisor
Dr. Kenneth A. McIsaac
The University of Western Ontario

Graduate Program in Electrical and Computer Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Duane Jacques 2013

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Recommended Citation

A Hybrid Visual Control Scheme to Assist the Visually Impaired with Guided Reaching Tasks

(Thesis format: Monograph)

by

Duane J. <u>Jacques</u>

Graduate Program
in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Engineering Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

In recent years, numerous researchers have been working towards adapting technology developed for robotic control to use in the creation of high-technology assistive devices for the visually impaired. These types of devices have been proven to help visually impaired people live with a greater degree of confidence and independence. However, most prior work has focused primarily on a single problem from mobile robotics, namely navigation in an unknown environment. In this work we address the issue of the design and performance of an assistive device application to aid the visually-impaired with a guided reaching task. The device follows an eye-in-hand, IBLM visual servoing configuration with a single camera and vibrotactile feedback to the user to direct guided tracking during the reaching task.

We present a model for the system that employs a hybrid control scheme based on a Discrete Event System (DES) approach. This approach avoids significant problems inherent in the competing classical control or conventional visual servoing models for upper limb movement found in the literature. The proposed hybrid model parameterizes the partitioning of the image state-space that produces a variable size targeting window for compensatory tracking in the reaching task. The partitioning is created through the positioning of hypersurface boundaries within the state space, which when crossed trigger events that cause DES-controller state transition that enable differing control laws. A set of metrics encompassing, accuracy (D), precision (θ_e), and overall tracking performance (ψ) are also proposed to quantity system performance so that the effect of parameter variations and alternate controller configurations can be compared.

To this end, a prototype called `aiReach` was constructed and experiments were conducted testing the functional use of the system and other supporting aspects of the system behaviour using participant volunteers. Results are presented validating the system design and demonstrating effective use of a two parameter partitioning scheme that utilizes a targeting window with additional hysteresis region to filtering perturbations due to natural proprioceptive limitations for precise control

Abstract

of upper limb movement. Results from the experiments show that accuracy performance increased with the use of the dual parameter hysteresis target window model $(0.91 \leq D \leq 1, \mu(D) = 0.9644, \sigma(D) = 0.0172)$ over the single parameter fixed window model $(0.82 \leq D \leq 0.98, \mu(D) = 0.9205, \sigma(D) = 0.0297)$ while the precision metric, θ_e, remained relatively unchanged. In addition, the overall tracking performance metric produces scores which correctly rank the performance of the guided reaching tasks form most difficult to easiest.

Acknowledgements

Personally, I would like to thank my family and loving girlfriend for their patience and emotional support through the years.

Professionally, I would like to thank my supervisor, Dr. Ken McIsaac for his input and guidance as I continuously hit roadblocks along the way. It is unlikely that a graduate student could find a kinder and more supportive supervisor. Without his help and the help of many other colleagues, I would have been at a loss. Some colleagues whom I'd like to notably thank are: Dr. Ranga Rodrigo for his collaboration during the early stages of this work; Akila Subasinghe for many fruitful discussions on the direction of this work; and Dr. Margret Davenport, for her advice relating to working with participants.

Lastly, I would like to thank some other friends and colleagues who helped make my time at Western an enjoyable experience: Dan Dechene, Kai Pisters, Paula Van Wyk, Elan Paulson, Ashley Warnock, Duncan Sutherland, Marco Luccini, Elena Uchiteleva, and Oscar Fillio.

Table of Contents

List of Tables

List of Figures

Acronyms and Symbols

CCD	*Charge Coupled Device*
DC	*Direct Current*
DEF	*Directional Error Fraction*
DES	*Discrete Event System*
DOF	*Degree of Freedom*
fps	*frames per second*
GLOH	*Gradient Location and Orientation Histogram*
GPS	*Global Positioning System*
HCI	*Human-Computer Interfacing*
IBLM	*Image-based Look and Move*
IBVS	*Image-based Direct Visual Servo*
ID	*Index of Difficulty*
MT	*Movement Time*
PBLM	*Position-based Look and Move*
PBVS	*Position-based Direct Visual Servo*
PC	*Personal Computer*
PPC	*Posterior Parietal Cortex*
PWM	*Pulse Width Modulation*
RMSE	*Root-Mean-Square Error*
RP	*Reaching Plan*
SIFT	*Scale Invariant Feature Transform*
SPL	*Superior Pariettal Lobe*
TMS	*Transcranial Magnetic Stimulation*
USB	*Universal Serial Bus*
$\tilde{S}$	*Set of DES-controller states*
$\tilde{X}$	*Set of DES-plant events*
$\tilde{R}$	*Set of DES-controller output symbols*
δ	*DES state transition function*

ϕ	*DES-controller output function*
$\tilde{s}_i$	*DES-controller state*
$\tilde{x}_i$	*non-silent DES-plant event*
ϵ	*silent DES-plant event*
$\tilde{r}_i$	*DES-controller output*
D	*State transition accuracy metric*
$\rho_x[n]$	*Directional Error Fraction*
θ_e	*Average DEF*
ψ	*Overall tracking response metric*

Chapter 1 Introduction

In recent years, a number of researchers have been working to adapt technology developed for robotic control to use in the creation of high-technology assistive devices for the visually impaired. These types of devices have been proven to help visually impaired people live with a greater degree of confidence and independence. However, most prior work has focused primarily on a single problem from mobile robotics, namely navigation in an unknown environment. The result of which is to either guide the user along an unobstructed path, or communicate the location of obstacles, and let the user determine their own path [2, 3]. The latter scheme accounts for the notion that in the world of mobile robotics, artificial intelligence is the fundamental limiting factor [4]. With current technology, an autonomous robot's ability to sense and process information about its environment far surpasses its decision-making capability.

In our previous work [5, 6] we presented a prototype assistive device aimed at providing an initial solution to the largely unaddressed problem of guiding a visually impaired person's hand to a target to complete a goal-oriented reaching task. The unit was a wearable assistive device which performed object tracking and visual servoing for a visually impaired user. The system captures images from a glove mounted camera, detects a given object of interest and directs the user's hand toward that target via a set of motion cues through a vibrotactile interface. That initial work is presented here, along with proposed further developments in terms of a model for the system using a supervisory hybrid control scheme. A significant issue that is addressed is the difficulty with measuring system performance given that from one user to another, and even across instances of usage by a given user, expected performance can vary greatly. So we extend the system model to incorporate a hybrid feature-space control scheme that provides a formalism allowing for the definition of new metrics that can consistently show real differences in reaching task performance for a nominal user.

1.1 Assistive Device Development in the Literature

Within the literature various researchers has attempted to address numerous aspects of improving independent living for the visually impaired community. One subset of locomotive navigational theme assistive devices constrain their framework to the recognition of text on signs within the environment (indoor or outdoor) to alert the user of conventional navigational markers used by sighted individuals position localization.

Work by Mattar *et al.* [7] and Silapachote *et al.* [8] involves detection, recognition, and identification[1] of text based signs within the environment for the purpose of improved mobility. Their system, named VIDI (Visual Integration and Dissemination of Information), acquires images from a head mounted camera unit. While sufficient for whole-body egocentric navigation, that choice of camera placement would be inefficient in the interface design for a reaching task as mapping to the reference frame of the hand would be exceedingly difficult.

Sudol *et al.* [9] proposed a system named LookTel that captures video from a mobile phone camera and streamed it to a desktop base station for feature extraction and object recognition. Object identifier tags were then sent to the mobile application which would vocalize the name of the recognized object to the user. With the assistance of human operator intervention at the base station, the mobile user could also request assistance with tasks such as identifying their current location and/or establishing a path to a destination, either directly or via waypoints.

Chen and Yuille [10] also proposed a client-server based architecture for text recognition on signs for urban navigation, but their work assumes that the visually impaired user is responsible for first aiming the camera at the text region of the sign and can take a steady still image (minimal blur) to supply the input for the recognition application.

Another large proportion of the assistive devices for the visually impaired are designed as path planning navigational aids. Systems such as the *GuideCane* [11] and the *NavBelt* [12]; use ultrasound, laser rangefinders, or stereoscopic camera rigs [13] to detect obstacles.

---

1. Text conveyed to the user through synthesized speech

Others, such as Coughlan and Manduchi [14] also proposed a similar navigational aid, but it is based on identification of strategically placed colored markers. This severely limits the usage of such a scheme in anything other than a controlled environment. The system developed by Hile *et al.* [15] chose to integrate GPS-data with their image based model for the construction of a navigation path for pedestrian wayfinding.

Zawrotny *et al.* [16] proposed a novel configuration for a haptic interface to feel the visual environment surrounding the user. Their system uses light-to-tactile transducer units mounted on the dorsal surface of each finger. The main transducer mechanism was comprised of a laser, solenoid, spring, and phototransistor. When a properly oriented edge is found through a change in the reflectance of the modulated laser beam, the mechanism begins to vibrate through oscillatory actuation of the solenoid against the spring mount. The tactile output would provide a constant 10Hz vibration signal as evidence of textural differences in the environment directly in line with the beam. The system did not aid the user in locating a desired object within the immediate environment. Arbitrary textural differences could not readily be identified as being edges of an obstacle in an open path versus opening in an oblique surface. That task was wholly dependant on the application of the user's intelligence.

In a subsequent publication [17] (from the same group of researchers as Zawrontny *et al.*) Stetten *et al.* , described an update that replaced the laser with a miniature camera and a vibrotactor. The system, now called FingerSight^TM, also evolved in purpose. They proposed that once an object is identified, gesturing with the finger can be used to enact remote control of the object. They describe an example of remotely flicking a light switch at a distance. However, no substantive details are given regarding which computer vision techniques are used for object identification, nor how finger movements would be detected and classified as command gestures.

Wanatabe *et al.* [18] present a "WEarable walKing" (WEK) camera based assistive system which provides dead reckoning through optical flow of edge features related to surface structures in the floor. They also attempt to delineate between wayfinding edges in the floor texture and those related to stairs or other architectural structures. The use of this type of design is predicated on some prior knowledge of floor surface textures in the intended indoor environment, limiting its use in unfamiliar indoor locations.

Yuan and Manduchi [19] presented a virtual "white cane" range sensing device

that employs active triangulation. Its use is modeled upon the characteristic pivoted sweep of a physical white cane. The authors use an Extended Kalman Filter to deal with the unknown, but approximately angular velocity of the user's sweeping motion. One of the key usability issues with this system is the need for training. The incremental planar scan depth and detection of obstacles is dependant on a consistent sweep pattern and rate.

A different white cane scheme, proposed by Kaneko *et al.* [20], for indoor environment navigation involved a system reliant on the installation of ceiling mounted beacons which broadcast localization codes. The user was notified of their current position within the map, via a receiver unit carrier on their person, as they passed beneath a beacon. Navigation from beacon to beacon way-points was accomplished through following a coloured line path marked on the floor. A colour sensor mounted on the "white cane" produced a vibrotactile output signal when the cane swept over the coloured guide line, providing an intermittent bearing signal. This, as with many other approaches in the literature require a significant level of augmentation to existing building or environment infrastructure which severely limits their likelihood of adoption.

Bigham *et al.* presented an assistive system called VizWiz in [21, 22] that heavily relies on remote human interaction. The system encompasses a broad scope of object and visual scene identification tasks through the data fusion of captured images and recorded verbal requests. The mobile phone application portion of the system captured images from the built-in camera and with verbal requests from the visually impaired user transmitted the data to a remote human assistant for interpretation within the context of the accompanying image. The requested tasks under investigation were primarily in the categories of *Identification*, *Description*, *Spatial*, *Reading*, and *Answering* [23]. Their solution (called VizWiz::LocateIt) to *Spatial* task bears some similarity to our proposed method in that they direct the hand-held camera to the target object via motion cues. Similar to our work, they advocate the use of a scale invariant feature transform (SIFT) based technique for objection recognition. In contrast, they provide motion cues to the user audibly through their "sonification" application module. The authors also do not report any detailed performance data about the usage of the system.

These types of mobile way-finding systems either incorporate the obstacles into a map of the environment so that an unobstructed path can be communicated to

the user, or they communicate the location of any immediate obstacles and let the user choose how to deviate from the current path to avoid the pending obstacle [2, 3]. The prevalence of the latter scheme is based on the notion that, within mobile robotics, artificial intelligence is the fundamental limiting factor [4]. With current technology, an autonomous robot's ability to sense and report far surpasses it ability to process information about its environment, driving its decision-making capability. As such, many researchers chose to leave the process of path planning to the much more versatile decision making capability of the human user. This approach is analogous to the cooperative interaction of a visually impaired person and their seeing-eye dog companion. In the partnership between dog and human, the dog provides navigation information and the human decides how to act upon that information.

Despite past research successes, adaptation of robotics and computer vision advancements to assistive devices lags behind the state of the art in mobile robotics research. One of the issues in this lag in research is the difficulty with incorporating the human user into models of performance and stability of such systems. The human, as the plant, within the control loop can make it very difficult to design a controller that provides a measurable, stable system performance. This fundamental issue will be addressed from multiple perspectives throughout the body of this work.

In this work we present a solution to an alternate, related robotics problem that is of great significance to the visually impaired; a goal-oriented, guided reaching task. The goal is very similar to the familiar robotics problem of servoing a robotic manipulator tool to a target object, but has been greatly under-investigated in the context of assistive devices. There is little reported in the literature, and the overwhelming majority of those assistive devices operate under the assumption that the user is somehow able to accurately aim a still camera at the target object. Or in the case of processing a video, that the user's visual survey (camera sweep) of the scene is systematic so that the camera alignment with the target object will automatically occur.

Drawing the obvious parallel of the human user's arm as the robotic manipulator, we approach the solution to this problem by creating a wearable assistive device that guides the user's hand to a given target object to complete the reaching task. There is some work reported in the literature, but with only moderate applicability.

One possible haptic solution to a visual servoing model of the problem could have an architecture similar to the exoskeleton type of force-feedback devices proposed

by various researchers (Bergamasco *et al.* [24]; Gupta and O'Malley [25]; Perry and Rosen [26]. These type of system essentially force the user's arm to follow a desired path for task completion. Consideration of this class of assistive device led us to pursue a different approach that we feel would be more efficient and considerably less cumbersome for the user. Assuming that the user's only impairment is with their sight, implementing a mechatronic system to physically drive their arm through a sequence of motions that they are amply capable on their is too intrusive. An exoskeletal apparatus approach could be highly valuable if the human user has some degree of motor control infirmity, but for user's without a motor impairment, the unnecessary weight and size of the required components restricts their freedom to interact with the environment instead of enhancing it.

Even though a guided reaching task is quite a different problem than the traditional locomotive navigation problem mentioned above, we can still use the motivating example of a seeing-eye guide dog to delineate the interaction between the user and system. Applying the analogy, we assert that the human retains supervisory control of the dog-human team, while the dog is only responsible for sensing the environment and passing that navigation information along to the human. In the same fashion, the proposed assistive device system identifies the valid target object and provides generalized guidance cues necessary to complete the guided reaching task. The user has the freedom to follow or ignore those cues at will. However, assuming the human chooses to follow those guidance cues provided by the dog, they are free do so with any arbitrary gait. In the context of a reaching task, the analogy translates to little or no limitation imposed by the assistive device on the pose and precise path that the user's arm follows during the act of the completing the reaching task.

The seeing-eye guide dog analogy also serves to differentiate our approach from another field of inquiry into assistive devices for the blind known as sensory substitution. The goal of most sensory substitution devices, (the classic example is the Optacon [27]) is to generate tactile cues to represent a scene pictorially. Patterns of raised and lowered pins attempt to give the user a tactile sense of the sampled scene. In essence, a haptic coding scheme for the image data, representing a visual scene in a manner the user can perceive via tactile input. Our approach, by contrast, can be considered the transmission of semantic information about a scene. Instead of attempting to represent a scene in a tactile format, we focus on conveying motion cues to the wearer. Some prior work by Tan and Pentland [28, 29] exists in the

field of sensory substitution that exploits the phenomenon of sensory saltation to give general direction cues to a user. Sensory saltation is achieved through vibratory stimulation of various cutaneous sites in sequence, the observer perceives the motion of the stimulus at interpolated points between the stimulus sights. While an interesting perceptual phenomenon, that technique may not be entirely suitable for the target tracking motor control cuing required for a guided reaching task.

1.2   Unique Application Considerations

There are three fundamental technical challenges in the system. The first problem is to design a visual servoing system that will recognize a desired object and generate motion cues towards it. By itself, this is a familiar problem from robotics research [30, 31], but with the added complication that the "robot arm" is now, in fact, a human arm. Since the system is human centric, it can not be easily proven to be a controllable closed-loop system. Within the loop, there can be significant issues with user's perception of control signals and their free will to choose whether or not to follow commands generated by the controller. Even under the assumption that the user makes a best effort to obey the control issued, their ability to do so can be affected by intrinsic physiological and psychological factors such as fatigue, frustration, or confusion. These are certainly not issues inherent in conventional robotic systems. However, for the purposes of this work we consider it a given that the user's intent is to follow the control to the best of their ability and thus refer to it as quasi closed-loop control system.

The second fundamental challenge is the development of a technique to communicate the necessary motion cues to the human user. Since the primary human sensory input channel (sight) is unavailable, we are forced to choose between lower bandwidth channels: namely audible input or haptic input. Using audible cues as a primary input channel is less desirable since the visually impaired already rely extensively on their hearing. Such additional audible input could be considered more akin to interference on the channel rather than a desired signal. Accordingly, the use of tactile cues to direct the motion of the user's arm was selected as the primary input channel to the user. The technique of using small forces to influence a user's direction has already been proven by other researchers in the field of assistive devices

(especially in the case of the *GuideCane*). While this tends to be used in assistive devices for navigation during a locomotion task, the evidence shows that users, visually impaired or otherwise, are very responsive to tactile cuing schemes.

The third challenge is tightly intertwined with the previous one. It is the design and/or application of relevant performance measures for this unique class of human motor-control application. As we discuss in further detail in Chapter 4: Human Motor Performance, various metrics for human motor performance can be found within the literature, but none of them fully address the distinct difference inherent in performing this type of task without normal visual input to the visio- and neuromotor systems of a human. Previous studies of human motor performance related to reaching tasks have been conducted with sighted individuals. Those studies initiated by Paul Fitts, brought about the development of Fitts' Law [32]. Through Fitts, and numerous other researchers, Fitts' Law has been used extensively in the evaluation of goal-oriented reaching tasks [33, 34, 35, 36, 37], but until recently its use has been with subjects that are sighted individuals. The lack of the normal human vision within the neuromotor control feedback loop significantly alters a user's motion planning capability. Without a usable set of performance measures, as with any control system, it is difficult to evaluate the behaviour of the system and then quantify the performance difference under alternate conditions.

1.3   Research Goal

The purpose of this research is to investigate the nature of guiding a visually impaired person's hand towards an object they wish to grasp through the use of an assistive device. From a basis of the natural process by which a person approaches a sighted grasping task, we have designed an assistive device system called `aiReach` (assistive image-based Reaching) to aid a visually impaired person in performing the initial stage the process, reaching for the object. To do so, a prototype assistive device was constructed as an experimental platform. An illustration of the hardware components that make up the prototype is given in Figure. 1.1(a), and a picture of the wearable components of the prototype is given in Figure. 1.1(b). Two versions of prototype were built during the course of this work. The first was an initial proof-of-concept to investigate general unknown usability issues. The second involved upgrades and revisions to the microcontroller and vision system software to enable proportional

control and increased frame rate, respectively. In our prototype the vision system
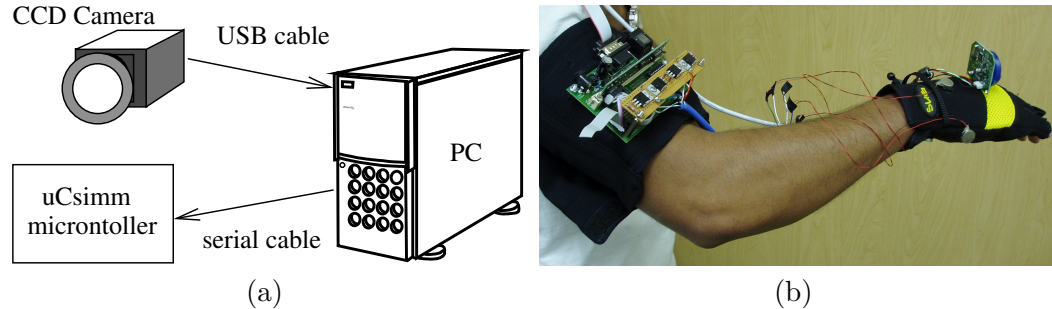
Figure 1.1: The `aiReach` (pronounced 'eye-Reach') system initial prototype: (a) illustration representing the three hardware components and their interconnection in the experimental prototype; (b) picture of the wearable portion of the system on the author's arm.

consists of a small, lightweight, colour CCD-camera and a PC workstation. The glove mounted camera is connected to the PC via a Universal Serial Bus (USB) cable. The glove is also equipped with four vibrating disc motors; one each on the palm, back of the hand, and either side. The placement of the four motors corresponds to the intended direction of motion of the user's hand that generates corresponding horizontal, vertical, or depth movements in the image plane. The microcontroller is connected to the PC via a RS232 serial cable and receives the trajectory data to general the appropriate motion cues.

The investigation includes analysis and a review of various computer vision techniques for object detection and tracking; using a Discrete Event System (DES) hybrid control approach to model the system; the development of a framework for measuring the performance of a goal-oriented reaching task for a non-sighted user; and a proposed motor performance model for this type of reaching task with the DES hybrid model.

1.3.1 Scope of the Work

Neither the construction of a commercially viable device, nor a prototype capable of normal operation outside of the laboratory environment is within the scope of this work. In either case the significant limiting factor is cost of system components. In particular, the cost of providing hardware acceleration to minimize the computation

time of image processing required for the object detection and tracking, and the power source (high capacity battery) for portability.

The type of limb movement involved is a constrained reaching task, as opposed to a full grasping task. The delineation being, that a reaching task would terminate at a point where the hand is sufficiently close to the target that grasping could occur with minimal probing of the immediate region of the task space. We do not address the problem of providing any pose control or force control law that can guide the user's hand to completion of the subsequent grasping task that would follow. This is considered a separate problem that is already naturally solvable through the user's precise motor control and sensory capability to probe the local area with their fingers to determine the appropriate pose and force necessary to grasp the object safely and securely.

Within the scope we present the design, modeling, and performance analysis of a prototype system used in experiments with voluntary subjects performing guided reaching tasks under an unsighted condition. The analysis is done on recorded trajectories from the numerous trials of guided reaching tasks performed during the experiments. The participants in each study are allowed and encouraged to respond to the motion cues in a manner natural and comfortable to them, so that the system response is indicative of realistic movement behaviours.

1.3.2   Research Contributions

Summarized below is a list of contributions we propose this work will make to the assistive device and HCI research communities.

- a proof of concept wearable assistive device to aid visually impaired users in a guided reaching task

- a feature-space hybrid control model for the system (non-sighted user - assistive device) behaviour during a reaching task using a somatosenory interface.

- demonstration using Fitts' Law that non-sighted reaching is not a simple ballistic pointing task.

- a proposed set of metrics to measure the guided reaching task performance that can quantify the effect of altering model parameters.

- a novel method for distance estimation with calibrated monocular vision using a non-uniform weighting scheme for relating object size to point-feature scale.

1.4   Organization of the Work

The remainder of this work is structured as follows: Chapter 2 gives an overview of the breakdown of the system. In that chapter we also provide the reader with some important insight in the unique considerations that make this problem very complex and how they motivated particular design considerations. Chapter 3 discusses the visual feature-space control law we have developed and the Discrete Event System (DES) model employed. Within that chapter, we address issues of visual feature extraction techniques and the need for feature-space based hybrid systems controller because of the absence of a known task-space and ground truth. The chapter outlines some trade offs between various feature tracking scheme investigated, and proposes a simple but novel technique for distance estimation employing a scale covariant and illumination invariant feature tracker.

Chapter 4 reviews the existing literature on human motor performance relating to reaching tasks and discusses the applicability of Fitts' Law, from the field of psycho-motor movement modeling, to measure the degree of success for goal-oriented reaching tasks. We also provide the development of new performance metrics based on the hybrid systems feature-space control model presented in the preceding chapter.

Chapter 5 describes the various experiments conducted to justify the material presented in this work. Conclusions and recommendations for Future Work are provided in Chapters 6 and 7, respectively.

Chapter 2 Describing the System

To create a system that addresses the fundamental aspects of the problem we must first define the context and scope of the operational task(s) the system must control. Continuing with the motivating analogy of the human and guide dog team, we describe in a general way the aspects of each system component.

In the context of a guide-dog pairing, the human receives motion cues for navigation conveyed through the dog's harness, but retains control of the dog/human team. The dog is only responsible for sensing the environment and passing navigation information along to the human through directional motion cues. The human, having supervisory control, still decides the speed, accuracy, and even whether or not to respond to the motion cues. As such we endeavor to create a wearable system that allows a human user to provide high-level planning while the assistive system generates motion cues based on visual sensory information. In response to a motion cue from the dog, the human is constrained in the way that they move. In a similar fashion, it would greatly increase usability across a spectrum of users if the system allowed for significant variation in arm postures during movement, allowing the user to determine what is comfortable for them during motion.

Within human motor control research field there are four core problems that drive the research [38]. They are the *degrees of freedom problem*, the *sequencing and timing problem*, the *perceptual-motor integration problem*, and the *learning problem*. All of these problems play some role in the development of the work presented here and warrant some explanation and consideration.

The human arm, not including fingers, provides a seven degree of freedom (DOF) manipulator for solution of a reaching task. In the conscious attempt to solve the reaching task, the human motor mechanisms will unconsciously produce kinematic and kinetic solutions that tend to be optimized for comfort. While the set of probable solutions will be similar to from person to person, each individual solution per identical reaching task can easily vary each time a repeated attempt is made for a given person. With some types of complex arm movements, there is a natural reduction in the number of degrees of freedom. For instance, in a pronated

or superanated orientation, when the elbow goes through flexion, so does the wrist. Having an opposing motion of extension at the wrist is possible, but requires conscious control. These natural coordinated motions can reduce the number of degrees of freedom that need to be considered for certain movement tasks.

In terms of a reaching task, the sequencing of movements is the main aspect of the second core problem to consider. A study done by Cohen and Rosenbaum [39] showed that coarticulation played a direct factor in the selection of point of contact when subjects had to grasp the handle of a tool, and it was dependant upon the next task in the sequence of movements. In the experiment, a toilet plunger was placed upright on a shelf of fixed height from the floor. Subjects were presented with a two task sequence, the first being to grasp the handle of the plunger and the second task was to retrieve and place the plunger on one of four shelves that were at different heights from the floor. Two of the secondary shelves were higher and the other two were lower than the initial shelf. The results of the experiment showed that vertical position along the handle where the user grasped was inversely proportional to the height of the secondary shelf, i.e. subjects grasped near the top of the handle if they were to move the plunger to the lowest shelf and grasped a low point on the handle when they were to place the plunger on a high shelf. The implication is that the sequence of tasks are presented as a coarticulation in the motion plan. Grasping a preferred point on the handle that minimized the necessary reach for the secondary task of moving the plunger to the target shelf was an optimal solution. Since the scope of this work is limited to only the initial reaching task, it is not necessary to have have apriori knowledge of subsequent tasks that could affect the determination of target point. It is enough for our solution to consistently aim for the center of mass of any target object.

The timing aspect of the second core problem does not require much consideration. In a single arm reaching task like this, the person is not attempting to coordinate more than one appendage, nor attempting to match some external rhythmic reference signal (dancing to music). Asynchronous coordination between motion cues and response movements are allowable and likely.

The perceptual-motor integration problem relates how perception affects motor control and vice versa. Perception and motor response are integrated together through mechanisms of feedback and feedforward control. Nearly all aiming movements proceed through two phases. They are initiated with a *ballistic phase* and

then followed by a subsequent *corrective phase* [40]. The *ballistic phase* exhibits an open-loop control behaviour while the *corrective phase* exhibits a closed-loop feedback behaviour. Additionally, some of the gross features of aiming movements incorporate a feedforward mechanism to integrate perception and motor response. The interaction of these mechanisms is not as well understood in the absence of normal human vision, and even less so in the presence of some form of visual sensory substitution.

The *learning problem* plays a crucial role in the acquisition and refinement of new and existing motor skills. It bears consideration in the development of assistive device research because it can affect the results of any system performance metrics that do not decouple the user behaviours from the control system behaviour. The literature [38] describes the learning problem in four different contexts. First is the process of *learning by doing* which helps the brain form correlations between the consequence of active motor commands and the resulting perceptual changes. Second is the *learning by deliberate practice* which is the frequent repetition of the skill with focused thought on the aspects of performance that require improvement. Ericsson *et al.* [41] showed that the amount of deliberate practice contributes more significantly to the development of a motor skill than the general hypothesis previously held within the field, that innate talent[1] was the most significant factor. *Learning through specific practice* is a similar, but distinct context. Precise specificity of practice of a particular skill will produce a narrow band performance increase. It is a key aid in the development of the kinesthetic sensations related to a particular set of movements. Keetch *et al.* [42] demonstrated evidence of this with an experiment conducted with basketball players shooting baskets at different distances from the hoop. Subjects made shots at varying distances, nine to twenty-one feet, from the hoop. With the exception of the foul line distance of fifteen feet, success rate percentage showed a linear relationship to distance for all other cases. The success rate was well above the predicted performance for attempts from the foul line because that was specific practice point common for all basketball players. Lastly, *learning through neural plasticity* is a context for motor skills development. It demonstrates the robustness of human motor control mechanisms in the presence of disturbance and the adaptation of performing a learned skill under a new set of environmental conditions; in essence, a capacity for generalization [38]. This property manifests itself in the reallocation of neural

---

1. From the viewpoint of performed motor skills such as playing an instrument or athletic skills

tissue that was once devoted to less frequently performed motor skill to one that is more frequently performed. Merzenich *et al.* [43] demonstrated this phenomenon through experiments with adult owl monkeys. They recorded neural activity in the somatosensory cortex before and after amputation of a monkey's middle finger. They observed that the segment of neurons that would fire in response to tactile stimuli of the middle finger gradually began to respond to stimuli at either the ring or index finger.

All of these learning contexts show evidence that there is an expectation that user of an assistive device will adapt to and form stronger correlative responses to the tactile guidance stimulus. This is obviously a desirable result in terms of the efficacy of the using the assistive technology, but has consequences in terms of being able to accurately measure the performance of the overall reaching task. Simple metrics such as time to completion, which are commonly used in the literature for assistive devices do not adequately separate performance improvements due to learning as opposed to improvements due to the controller design.

2.1 Reaching and Visuomotor Trajectory Planning

Research into the processes involved in human motor control related to reaching tasks has shown that there is a significant amount of flexibility in the mapping of sensory information, perception, and action between the human neuromotor and vision systems. Two distinct components of the human visual system are utilized in the process of aiming: object identification for determination of the target and motor path planning via visual guidance. Execution of aiming for the reaching task is accomplished through the coordination of a preprogrammed feedforward control subtask and a error correction feedback subtask. Target selection, the initial movement plan, and continual updates to the trajectory have been shown to be controlled by a combined contribution from the posterior parietal cortex (PPC) and superior parietal lobe (SPL) [44, 45, 46]. Desmurgert *et al.* [47], in a study involving transcranial magnetic stimulation (TMS) of subjects' contralateral PPC during reaching tasks, were able to disable the corrective phase of the movement. Their results indicated that the initial ballistic (feedforward) phase did not require parietal control.

But when exerted, parietal feedback based corrective movements could override the current movement in progress. TMS pulses applied when the ipsilateral hand was performing the reaching task seemed to have no effect. This is a further indication that the PPC does not act in a solely visual role [44]. Work by Stuphorn [48, 49] and colleagues showed that there are two groupings of neurons in the superior colliculus (SC) that respond during arm movements. The indications are that one grouping corresponds to the use of a gaze-centered reference frame and the other corresponds to an intrinsic representation of the movement that is either in a muscle and/or joint reference frame.

A study by Gordan and Ghez [50] on planar accuracy in aiming showed that the spread of reaching task terminal points was elliptical about the target. The major axis of the elliptical spread was along the target line connecting the initial point and the target. The minor axis fell along lines perpendicular to the target line. Those results imply that the subjects demonstrated a greater accuracy in direction over accuracy in amplitude. This result also correlates with ballistic phase of a reaching movement. Thus during the corrective phase, amplitude inaccuracy requires either a correction for overshoot or for undershoot. Gordon and Ghez also found [50] that for both sets of the near and far targets presented to the subjects, the velocity profiles were bell-shaped. Their results indicate that the motor planning mechanism for the aiming movement relies on a preprogrammed feedforward control scheme for both distance and direction parameters before the onset of the movement. Prior to that, Vince and Welford [51] had published results that showed that correction for overshoot is more costly in time and energy. Some of the earlier work by Rosenbaum [52] showed that with the reaction time to onset of an aiming movement, subjects took more time to resolve direction uncertainty than amplitude uncertainty. This could account for why magnitude of amplitude errors are greater then that of directional errors.

The above studies and many others involved subject with a full clear view of the targets to which they were aiming. The earliest work that involved the targeting accuracy of aiming movements with subjects under a non-vision condition was performed by Woodworth [53]. Subjects were asked to perform reciprocal target tapping movements in time with a metronome. The experiment was performed under two visual conditions. In one the set of trials the subjects had their eyes open and in the other their eyes were closed. Woodworth found that under the eye-closed condition a subject's movements were entirely preprogrammed and ballistic in nature. As ex-

pected, under the eyes-open condition a subject's movements were comprised of both ballistic and corrective movements. The experiment also produced an interesting result when he plotted the movement velocity against the mean absolute targeting error. For the eyes-closed condition, the mean absolute targeting error was approximately constant across the spread of movement velocities (5cm/s to 50cm/s). However, plotting the same for the eyes-open condition showed that the mean absolute targeting error increases monotonically as movement velocity increases, and converges to the same value as the eyes-closed condition trials (approximately 4.5mm). From this result, Woodworth postulated that if the movement time was short enough so that the movement was completed before the corrective phase was initiated, then the targeting error would be the same regardless of whether the subject's eyes were open or closed. Thus visual feedback would no longer effectively increase targeting accuracy. Woodworth calculated that movement time threshold at approximately 200ms. Keele and Posner [54] also produced a similar estimate of 200ms for the movement time threshold before visual feedback can be initiated. Subsequent studies by Carlton [55], and later Zelaznik *et al.* [56] derived estimates as low as 100ms. In any case, an upper bound of 200ms is accepted within the literature. The material presented in the previous sections clearly shows that under normal visual condition the human user performing a reaching task can be modeled as system with both feedforward and feedback control mechanisms. With normal human vision, the combined application of both mechanisms form the initial stages of the reaching task planning and provide the capability to precisely control the complex multi-joint dynamics required to complete the movement [57]. The internal model that represents this control law requires both intrinsic (proprioception and kinesthesia) and extrinsic (perceptual and task specific) information to form the necessary motor plan. A block diagram of the combined feedfoward/feedback control for a reaching task, as proposed by Jordan and Rumelhart [1], is depicted in Figure 2.1.

The reaching plan, RP, is formulated and updated via the difference between the desired hand position, $\bar{\mathbf{x}}$, and the estimated position of the hand, $\hat{\mathbf{x}} = f(\hat{\mathbf{x}}_{\mathbf{m}}, \hat{\mathbf{x}}_{\mathbf{s}})$. The desired control, $\mathbf{u}$, drives the arm in an attempt to match the plan, resulting in an actual hand position, $\mathbf{x}$. The terms $\hat{\mathbf{x}}_{\mathbf{m}}$ and $\hat{\mathbf{x}}_{\mathbf{s}}$ are the contributory estimates of hand position from the forward (predictive) model and from sensory feedback, respectively [1, 58, 59].

Several researchers have investigated whether the reaching plans are executed
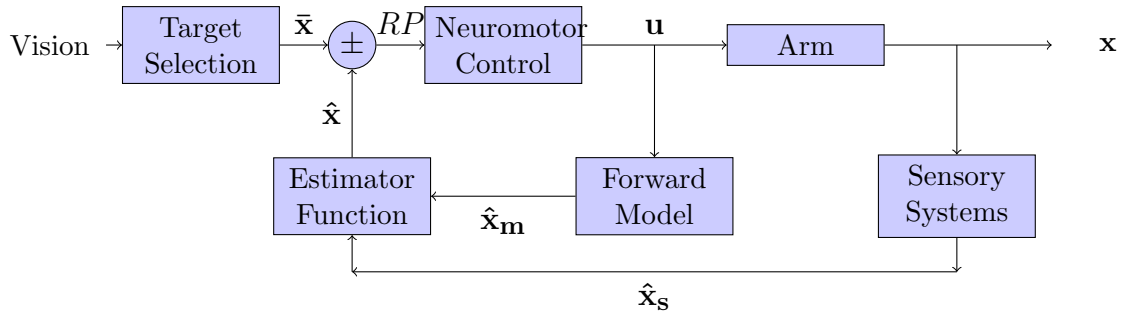
Figure 2.1: Block diagram of an internal feedforward/feedback model [1] with arm state estimation for compensatory tracking in a visually guided reaching task.

in joint-space using intrinsic body coordinate frames or in hand-space with extrinsic coordinate frames. In a series of two-dimensional surface pointing tasks, Morasso *et al.* [60, 61] found that subject's hands tended to move in straight line segments, even when asked to draw curved paths. In contrast, Soechting *et al.* [62] found in a set of experiments where subjects were asked to point from an initial rested, dangling arm position to a target in a vertical plane in front of them, that the peak angular velocities of the shoulder and elbow coincided. The results also showed that the ratio of peak velocities of the two joints was equal to the ratio of radial distances that joints moved through. Soechting *et al.* surmised that this relationship of ratios was evidence that some aspect of the joint-space was taken into account during motion planning. The literature is unclear as to how or why a hand-space or joint-space solution would take precedence in the reaching plan, and that it is an area for further investigation.

2.2   Vibrotactile Interfaces

The system is equipped with vibrotactile output transducers to convey the motion cues to the user. Various studies have been published in the field of Human-Computer Interfacing (HCI) that try to determine characteristics that significantly affect the utility of vibrotactile interfaces. A large segment of this work exists in the context of teleoperation of robotic systems. Some works have produced guidelines for vibrotactile interfaces, but until recently they were primarily concerned with passive displays such as Braille labels on keyboards and other control interfaces for telecommunica-

tions devices [63]. That is not to say that there isn't a wealth of different vibrotactile interface design reported in the literature. In fact, there are too many to list comprehensively. However, until recently [64, 65], very little had been published that compares and contrasts the existing work in the field to formulate a coherent set of guidelines for use of existing findings.

Verrillo [66, 67, 68] produced some of the earliest studies that showed that there are four key parameters to consider in vibrotactile interface design: amplitude, frequency, timing, and placement. The parameters, individually and in combination, can greatly affect the effectiveness of a tactile interface through comfort, stimulus detection, and tactile information coding. Design consideration within each of the categories must make allowance for the high variability of conformance from user to user, so an ideally designed interface should allow for tuning. As an example, Sherrick and Cholewiak [69] found that skin is roughly sensitive to vibration in the range of 20-250Hz, with a minimal amplitude detection threshold of 4 microns at 200Hz along surfaces on the torso, but that spatial and temporal acuity of detection can greatly degrade with aging.

Comfort is an aspect that is strongly user specific. However, Van Erp [64] did report some generalized guidelines for comfort thresholds. The author states that care should be taken to minimize heat transfer from the vibrotactor to the skin, amplitudes above 0.6-0.8mm can elicit a pain sensation, and that the musculoskeletal structure of the hand-arm is more susceptible to injury from extended exposure to signals of approximately 12Hz.

2.2.1 Stimulus Detection

In terms of amplitude, frequency, timing, and placement parameters, not all body parts have the same degree of sensation acuity. Glabrous skin is more sensitive than hairy skin, particularly when it come to stimulus localization. Even across the various hairy skin surfaces of the body such as the arms, face, and trunk, the minimal spacing for a locus of stimulus points can differ substantially. The highest sensitivity for stimulation detection occurs across the frequency range of 20-250Hz.

The temporal sensitivity of skin is very responsive, but the psychophysical phenomenon of temporal summation can lower the detection thresholds. Early studies by Gescheider [70] and then later by Pestrosino and Fucci [71] showed that detection

was possible at bursts produced by 10ms pulses with 10ms gaps between. However, if the frequency is ramped, a smoothing effect can occur which perceptually smothers the temporal sensitivity, thus requiring a slower pattern for coherent detection. This has implication towards coding and achievable tactile symbol rate.

Stimulus detection can also be negatively affected by temporal masking. Somatosensory receptors tend to integrate prolonged vibratory stimulus so that the waveform of frequencies applied can have an effect on stimulus detection thresholds. There is decreasing sensitivity to stimulus waveform patterns in order of: square, triangular, and sinusoidal [64]

2.2.2 Tactile Information Coding

Similar to stimulus detection, the four parameters of amplitude, frequency, timing, and location of vibratory signal affects the way and depth information is coded as tactile symbols. Beyond a basic binary (on/off) alert messaging scheme, tactile symbols can be coded via subjective detection of signal magnitude (intensity). However, subjective magnitude is a non-linear function of the applied amplitude and frequency. Early work by Craig [72] had suggested that no more than four different levels of intensity between detection threshold and comfort threshold should be used to ensure accurate discernibility of symbols. That followed research by Goff [73] that suggested coding via frequency variation should not exceed nine different levels, and the difference between levels should be at least 20 % from the adjacent level. Mutli-element interfaces can be extend the size of symbol set through coding via location of the stimulus. A important consideration with this technique is the actuator density of the interface. Certain regions such as the fingers, hands, and face can accommodate dense multi-element arrays of actuators. In some early work by Johnson and Phillips [74], the authors made the claim that when a minimum spacial acuity of 4mm was acceptable, any locus pattern was acceptable for untrained users. However, a higher resolution could be achieved with the addition of training. Their experiments dealt with spatial acuity on the glabrous skin of the hand which is known to have high mechanoreceptive sensitivity. Cholewiak and Collins [75] tested localization accuracy around the waist using straps with various sets of equidistantly space vibrotactors. The three straps were equipped with 12, 8, and 6 vibrotactors at 72mm, 107mm, 140mm inter-actuator spacing, respectively. The results reported average lo-

calization accuracy rates of 74%, 92%, and 97%, respectively. The ill combination of certain design choices of location, frequency, and timing parameters can generate unintended spatial effects in decerning tactile symbols. When simple pattern based coding schemes are used for symbols and there is an overlap in the timing of neighbouring stimulation sites, spatial masking can occur. It can cause the user to perceive a single apparent location for the stimulus between the neighbouring sites [76]. This perceptual phenomenon can be avoided if distinctly different frequencies, during temporal overlap, are used at the neighbouring stimulation sites [77]. If the use of apparent locations is intended, the spatial masking can be enhanced by ensuring that the stimulus frequencies are matched and in phase. This would require a precise actuator tool (contact surface) with a fixed home position. This is often achieved in the literature using rectilinear vibrotactors that act normal to the skin surface, as opposed to the tangential forces induced by disc motor with eccentric weighting on the rotor.

It is desirable to design an interface with minimal complexity between the symbols conveyed and their corresponding motion cues so that user does not have to expend much cognitive effort to track the intended trajectory. It has been documented that the performance of motor tasks can suffer when a person focuses too much attention on the precision of their movements [78]. Simplicity can be accomplished through orthogonal basis signals which maximize the discernibility of the symbol's signal components . Perceptually orthogonal signals can be produced through a number of methods including spatial location of the stimuli, frequency of vibration, and burst rate (vibration pattern).

Phong Pham and Chellali [79] investigated the tactile cue signalling performance in the context of a vibrotactile interface for teleoperated robotic control. The purpose was to investigate the mapping between the robot's sensory-motor space and the human user's sensory-motor space.

2.3 Modeling a Visually Impaired User as the Plant

Conventionally, the initial step in designing a control system is to model the dynamics of the plant. There are numerous proposed models in the neuromotor movement

modeling literature. The literature provides evidence of that the human neuromotor system is easily capable of solving the inverse dynamics or kinematics for an extreme large scope of complex arm movement tasks. The human neuromotor system can almost effortlessly solve the inverse dynamics problem of a typical reaching task. Regardless of the various proposed models, the focus within this section is to describe the high-level issues that relate to designing a controller for a human plant.

The key notion that must be clear is that we wish to fully leverage the physical acuity of the user and issue motion cues that specify a stable trajectory between the reaching task terminal points. This should includes allowance for user comfort during the movement task.

The obvious initial inclination is to model the system as a visual servoing problem, but there are some natural constraints that must be taken into account. Even though the human arm can be viewed as a +6-DOF manipulator with differing sets of holonomic constraints on the various joints, there are some natural movement tendencies the allow for a significant reduction in the number of DOFs for object tracking. As a simple example, consider the lateral flexion and extension of the wrist. While it is possible to laterally move the wrist through a moderate extension arc and a very limited flexion arc, in a pronated position the wrist is normally axially aligned with the forearm because it requires minimal exertion. Given this natural movement behavior, we can reduce the problem by 1-DOF under the reasonable assumption that the user will maintain a wrist orientation that has minimal deflection from the forearm axis. Similar cases can be made for further reductions.

Regardless, we still find it valuable to examine the system in the context of a visual servoing problem. Using the taxonomy introduced by Sanderson and Weiss [80] which classifies visual servoing systems into four main categories: *Dynamic Position-based Look and Move* (PBLM), *Dynamic Image-based Look and Move* (IBLM), *Position-based Direct Visual Servo* (PBVS), *Image-based Direct Visual Servo* (IBVS); this type of visually guided reaching task is considered to be an IBLM system. It is categorized as such because the feature-space controller uses point features extracted from each image to generate the set-points for the joint-space controller (the user) even though that error signal is in terms of camera motion, in feature-space, as opposed joint-space variables. The IBLM model is illustrated in Figure 2.2.

Thus we can view the overall system in layers, where the plant for the assistive device control system is the human user, but that human plant is viewed as servoing
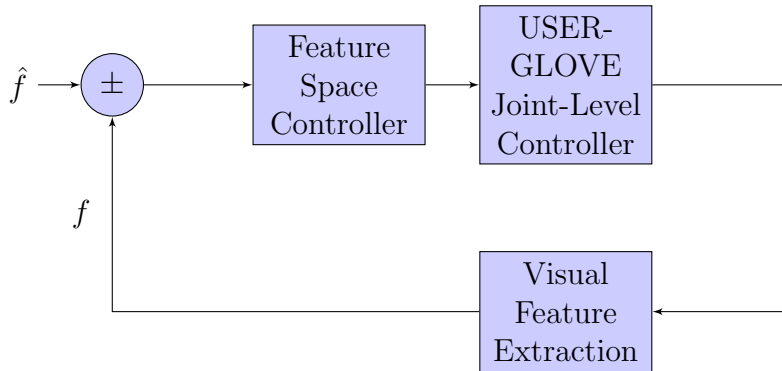
Figure 2.2: A block diagram modeling the assistive application at an high level abstraction as a generic Image-Base Look and Move visual servoing system

manipulator with an unconventional joint-space control scheme. In this way, the block diagram given in Figure 2.1 is contained within the USER-GLOVE Joint-Level Controller block of Figure 2.2.

The system uses a monocular vision eye-in-hand configuration and the user's hand is considered to be the end-effector. Eye-in-hand systems are said to be endpoint open-loop because the system only observes the target object, while visual servoing systems that employ a camera at a distance to the end-effector are said to be end-point closed-loop as both the target object and end-effector are seen [31]. Without an external reference camera, precise contact registration between the endeffector and the target cannot necessarily be achieved. An external reference view is generally desirable for a servoing task, but for a wearable assistive device it can be untenable. There are a number of issues with providing a second camera view. It is cumbersome and impractical to rigidly mount a second camera to some other part of the user's body that can clearly provide an external reference view of the end-effector proximity to the target. Providing a second camera view entirely external to the user is counter productive as it constrains the use of the assistive device to only that locale. However, we can leverage the user's intelligence and fine motor control ability in place of a number of key control processes. If we redefine the reaching task target position as some region in task-space that is very close to, but not in contact with the target object, then the goal is get the end-effector *close enough*. Once the endeffector reaches a point within that target region, the system can "transfer" control to the user to probe for the physical object; determine the appropriate orientations and forces for the tool (fingers) necessary to grasp the physical object.

Generally, an image based visual servoing system follows an proportional error control law that compensates for the difference between the goal-view feature vector $\hat{\mathbf{f}}$ and the current-view feature vector, $\mathbf{f}$, through camera motion given by

$$\dot{\mathbf{p}} = \mathtt{K} \cdot \mathtt{J}_v^+(\mathbf{p})(\hat{\mathbf{f}} - \mathbf{f}) \tag{2.1}$$

where $\mathtt{K}$ is a gain matrix, $\mathtt{J}_v^+$ is the pseudo-inverse of the image Jacobian, and $\dot{\mathbf{p}}$ is the velocity screw of the terminal point on the end-effector. It can be advantageous to use moments of image feature within the feature vector as IBLM visual servoing from point features alone can result in infeasible camera motion due to the coupling between translational and rotational degrees of freedom in orientation errors and trajectory traversal [81].

In a conventional visual servoing application we have known, consistent, and precise physical dimensions for the manipulator, focal length of the camera, etc. So assignment of various coordinate frames; tool, camera, joint(s), base, and world are made. Thus the velocity screw can be defined as

$$\dot{\mathbf{p}} = \begin{bmatrix} T_x \ T_y \ T_z \ \omega_\alpha \ \omega_\beta \ \omega_\gamma \end{bmatrix}^T \tag{2.2}$$

with T_x, T_y, T_z denoting the translation velocities of the end-effector terminal point with respect to the manipulator's base frame and ω_α, ω_β, ω_γ denoting the rotational velocities about the base frame's X-, Y-, Z-axis, respectively. In this kind of wearable application there is no consistency of physical parameters from one manipulator unit to another (different users). In essence the only physical parameters that are consistent are the camera related parameters: resolution, focal length, angle of view, etc. Hence we can only rely on a consistent image-space.

The transforms between the base and various joint frames are not known, but as everyday experience tells us, they are not required for a movement solution. The transforms are part of our internalized neuromotor representation for the movement. Given that we are working with a visually impaired person in an unmapped environment, the task space is not fixed. It can extend beyond the arms length to any distance at which the target object is identifiable. Since visually impaired persons' perceptual frame of reference is themselves, the task space is anchored by a egocentric frame of reference and can easily move within the external environment.

Since the users themselves inherently act as the robust joint-space controller for

the manipulator (arm), all torque control is also handled by the user, and the end-effector (hand), during natural movements, moves in hand-space coordinate frame, we can view the problem as control of 6-DOF kinematic point at end-effector.

In previous sections reductions in degrees of freedom related to natural coarticulations and sequencing of movement tasks, but we can also employ functional reductions in the number of degrees to simply the problem. Consider the functional task space for the reaching task. There is no need for the system to guide the user's hand toward their own body. Thus postures of the arm-hand that are not directed outward from the body are unnecessary. Therefore in the context of the reaching task, the plant in this problem should be viewed as a 3-DOF kinematic point.

2.4 Designing a Controller for a Human Plant

It is an extraordinarily difficult task to create a system that can provide the high capability of a human to identify a target, localize the relative position of the hand, localize the terminal points of the reaching task, and calculate the entire trajectory of the movement in real-time.

While the neuromotor system as a joint-space controller is very robust in the presence of disturbances, its accuracy is not deterministic. The user's cognitive ability is highly robust, but the stability of the "actuators" and joint "sensors" (neuromotor muscle activation, kinesthesia, and proprioception) can be susceptible to physiological and psychological factors. For example, a conventional robotic manipulator becomes neither tired nor frustrated.

Consequently it is important to first determine what are the crucial state variables and what aspects of the system need to be controlled. It is beyond the scope of this work to create a system the can incorporate a sequence of additional motor tasks after the reach itself. Thus, we can ignore the coarticulation issue of the timing and sequencing problem and specify the terminal point of the reaching task as being in the immediate vicinity of the target object. Consequently, the controller does not need to determine a preferred contact point on the target to mimic the natural behaviour for sequenced tasks exhibited in Cohen's and Rosenbaum's [39] study, described earlier. As such, it is reasonable to specify the centroid of the projected surface as the optimal point to target for contact proximity to the object.

In the previous section we argued the merits of modeling the system as a 3-DOF kinematic point. Proceeding upon that notion, we must specify what level of kinematic control is achievable through a tactile interface.

For the design of the aiReach (pronounced eye-Reach) system we have chosen to codify the motion cues issued to guide the hand as a series of mechanical vibration signals conveyed through somatosensation; creating a vibrotactile interface from the aiReach controller to the plant (user). Since we wish to maximize tactile symbol
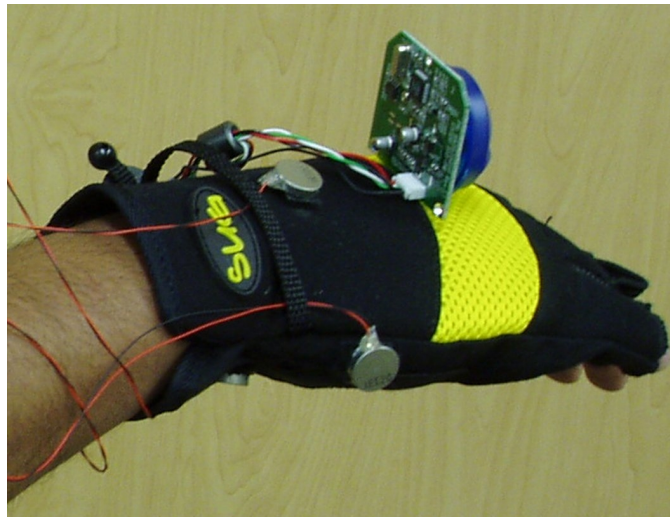
Figure 2.3: Image shows the glove portion of the aiReach prototype to illustrate the placement of vibrotactors and camera. The dorsal and ulnar ipsilateral vibrotactors are visible in this right-handed configuration.

discernibility, we have chosen an orthogonal signal locus for the interface, consisting of four small variable-input DC vibrotactors (eccentric imbalanced disc motors). The vibrotactors are mounted on a fingerless glove in pairs on the lateral, dorsal, and palmar parts of the hand. These orthogonal pairs form a planar axis orthogonal to the medial metacarpal bone and the plane is approximately parallel to the image plane of the camera mounted on the glove distally ahead of the vibrotactors. They are clearly visible in the close up picture of the aiReach glove shown in Figure. 2.3. The individual vibrotactors are used to convey a set of symbols relating motion cues to the user. The language consists of six symbols indicating the following motion cues: `Move-Right`, `Move-Left`, `Move-Up`, `Move-Down`, `Move-Forward`, and `Stop`. There is no need for a `Move-Backward` symbol as that implies that the camera is

past the target, which also implies that the target is not visible. Quite simply, if the target is not visible the required motion cue to acquire it is indeterminate.

The four vibrotactors, individually actuated, convey intended motion in two degrees of freedom within image space. The `Move-Forward` and `Stop` symbols are conveyed through simultaneous actuation of all four in different patterns. Continuous actuation indicates a `Move-Forward` cue, and a pattern of four uniform intervals of pulses indicate the motion cue of `Stop`. It should be noted that the mapping of the dorsal surface of the hand to the direction up is based on an imposed convention for the pose of the hand; that it is in a pronated orientation. For the remainder of this work, we will adhere to that convention as the nominal pose for the hand. It it worthwhile to mention that by defining the motion cues in this way, in the direction of the surface normal at those points on the hand, a correct trajectory to the target object can still be followed regardless of the whether the hand is in a pronated, supinated, or neutral orientation because the transformation from the camera frame to the image frame to the hand frame is fixed and the user's proprioceptive capability makes it a trivial task to intuit the transform from hand frame to egocentric frame.

The technique of resolving the error in each axis separately was selected because of the difficulty in issuing a motion cue in polar form: $(|\mathbf{r}|, \theta, \phi)$ to the user through a glove mounted vibrotactile interface. Responding to a signal of magnitude $|\mathbf{r}|$, as a motion cue, is no more difficult to interpret than a magnitude in one of the axial directions, but interpreting its direction consistently would be problematic across various subjects. The problem lies in the subject's ability to interpret and track a precise values of both θ and ϕ. While it is possible to issue values of $\theta = n\frac{\pi}{4}$, with $n \in \{0, \ldots, 7\}$ using the four vibrotactile outputs employed, the visually impaired user does not have access to the feedback necessary to confirm (closed-loop) that their hand is precisely following a trajectory along that initial directional component. The problem is even more pronounced with the other directional component, ϕ, out from the image plane. The feedback, which we take for granted, is the ability to see the motion of our hand in reference to the horizon. Without that visual horizon reference, a user would need to have a precise kinesthetic perception and substantial training and practice to effectively track the motion cues indicating trajectories other than horizontal, vertical, or frontal.

Furthermore, taking in account the difficulty for the user to resolve any angular trajectory motion cues it becomes apparent that trajectory generation must be calcu-

lated on an axis by axis basis per image, in some sequential manner. The necessity of sequencing a trajectory traversal between terminal points of the reach, or "tracking" state on an axis by axis basis leads us to adopt state automata based scheme for our control. The details of that model and a detailed discussion of the hybrid control driven by feature-space parameters is given in Chapter 3.

It should be noted that no force control scheme for grasping the target object is necessary. The goal of the controller is to guide the user's hand to within a reasonable distance just short of collision with the target object. Then the user determines[2] the appropriate end-effector pose and level of force used to grasp the target object.

Another crucial application consideration stems from the perspective of a human plant is a lack of precise position control of the end-effector (hand). Since it is impossible for a user to hold their hand perfectly steady so that the centroid of the target projection and the center of the image frame align with precise registration, we use a broader definition of being "On-Target". This is accomplished by mapping the precise target position, $\mathbf{p}(t) = [p_x \ p_y]^T$, to an approximate target error, $\mathbf{v}(t) = [v_x \ v_y]^T$, within the image frame. The mapping produces a variable sized targeting region within image-space given by equation (2.3).

$$\mathbf{v}(t) = n(l)\mathbf{p}(t) \tag{2.3}$$

The scalar function $n(l)$, given in equation (2.4), creates a deadband region, presuming that as long as the centroid of target object projection, $\mathbf{p}(t)$ is within $\pm lI_x$ away from the horizontal center of the image and/or $\pm lI_y$ away from the vertical center of the image, the user is considered to be "close enough" to On-Target. The parameters I_x and I_y are the horizontal and vertical image dimensions in pixels, respectively.

---

2. Through systematically feeling the shape, orientation, texture, etc of the object to determine the best way to grasp it.
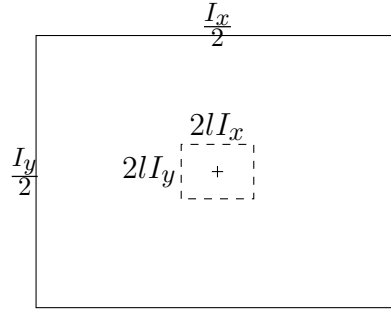
Figure 2.4: An illustration of the approximate target region within image space
defined by l. The region dimensions are $2lI_x$ by $2lI_y$

The `On-Target` region, centered at $[\frac{I_x}{2}; \frac{I_y}{2}]^T$ is illustrated in Figure (2.4).

$$
n(l) = \begin{cases}
1 & \text{if} \quad 0 < o_x(t) < \frac{I_x}{2} - l\ I_x, \\
& \qquad 0 < o_y(t) < \frac{I_y}{2} - I_2 \\
0 & \text{if} \quad I_x(\frac{1}{2} - l) \leq o_x(t) \leq I_x(\frac{1}{2} + l), \\
& \qquad I_y(\frac{1}{2} - l) \leq o_y(t) \leq I_y(\frac{1}{2} + l) \\
1 & \text{if} \quad I_x(\frac{1}{2} + l) < o_x(t) < I_x, \\
& \qquad I_y(\frac{1}{2} + l) < o_y(t) < I_y
\end{cases}
\tag{2.4}
$$

We define the value of l such that $0 \geq l \geq \frac{1}{2}$ so that it corresponds to a percentage of an image-space dimension. Thus lI_x, a given number of pixels, translates into the physical size of the target region plane in task-space at a particular distance from the real target object. This allows `on-target` region in task-space to vary hyperbolically with distance from the camera[3] to the real target object. Thus, when the user's hand is a moderate distance from the real target the registration accuracy is quite broad, but narrows as the camera approaches the real object in depth along the optical axis.

The target object location is only resolved in image space using the feature set parameters describe in Chapter 3. Regardless of the feature extraction technique employed, all that is required is the calculation of the moment based feature that represents the centroid of the object projection in the image. The image-space coordinates for the centroid of the target object are given by $\mathbf{p}(t)$

---

3. Assuming the use of a pinhole camera model

Once the target object is identified, the displacement of its centroid from the center of the image is used as a measure for generating control signals which gives the motion cues to the user's hand. The same procedure is followed with each subsequent image, thus creating a *quasi-closed loop*. We use the term *quasi-closed loop* to represent the idea that whereas we are generating the appropriate control signals for compensatory tracking, the plant (user) is free to ignore them. The control signals are resolved as two components: image space horizontal and vertical offset from the center of image and the distance from the camera frame to the surface plane of the of the object normal to the optical axis of the camera. The controller input $\mathbf{e}_i = \hat{\mathbf{f}} - \mathbf{f}$, where $\hat{\mathbf{f}} = \left[\frac{I_x}{2} \; \frac{I_y}{2} \; f_\zeta\right]^T$ and $\mathbf{f} = \left[p_x \; p_y \; f_z\right]^T$. The depth estimations of f_ζ and f_z are given by some moment generating function, $g(f_i), i = 1, \ldots, n$, of the image features used to identify the object. The function is specific to the particular set of features extracted based on the specific computer vision technique used. In Chapter 3 we describe the three techniques used in this work and describe the corresponding $g(f_i)$.

2.4.1   Image Acquisition and Processing

Only a few basic criteria were used in the selection of a camera for the system. Ideally, it should be an inexpensive, light-weight, off-the-shelf camera with acceptable image quality and capable of full motion video frame rate. Two classes of cameras were initially investigated as possible candidates. The first was a camera capable of generating images with a resolution of 1024x768 at 30 frames per second (fps) with IEEE1394 connectivity. The second was a webcam capable of generating images with a resolution of 320x240 at 30 fps and transferred to the PC via a USB 2.0 interface. Some issues related to the comparative performance results for the two cameras are discussed in Chapter 5

Another key criteria that has been chosen is that the system will only employ monocular vision for sensing. This choice was made due to a number of factors that were a natural consequence of the application being a wearable assistive device, some of which are: minimizing weight, size, cost, and obtrusiveness of the wearable device; allowing for unique physical dimensions of each user; and that the pose of the end-effector (hand) is not controllable.

Intuitively, the choice of monocular vision might seem to be irregular as ranging data is necessary for determining the depth movement into the task space towards

the target. Two of the typical methods for obtaining range data for an eye-in-hand configuration when visually servoing towards a target within an undefined task space are the use of stereoscopic vision, or a laser range finder in conjunction with a single camera. Stereoscopic vision rigs suffer from the constraints in the first two factors listed above and the effectiveness of a laser range finder is severely hampered by the third factor. Using simple computer vision techniques for approximating object size with a monocular vision configuration provides sufficient ranging accuracy to achieve the broader performance criteria of "close enough".

2.5   System Performance Criteria

As with any control system it is essential to develop and describe metrics characterizing the performance of the system. We have proposed that the system can be decoupled into a coordinated set of one-dimensional problems, each with simple a state equation, $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$. However, as it is hopefully becoming clear, the plant (human user) is not actually controllable. Thus, any performance metrics must incorporate the notion that measuring the responsiveness of the system includes the variability of contributing factors in the inherent physical acuity of a particular subject and expected performance of the system in relation to the all potential subjects.

Thus, instead of a continuous performance metric we propose a discrete set of performance levels that quantize a range of continuous values into a behavioural class. These proposed, generalized classes are:

Ideal indicates that user followed the motion cues with a high level of precision and accuracy;

Well-behaved indicates that the user had followed the motion cue with a reasonable degree of precision and accuracy;

Ill-behaved indicates that the user exhibited, cumulatively over the entire trajectory, a significant deviation from the motion cues issued.

The definitions of the three classes are given in vague terms at this point because the description of our proposed hybrid control systems model based on a finite automata is not provided until midway through Chapter 3. The material presented

in those sections is necessary before further detail definition of the behavioural class can be discussed.

We also present a description of human motor performance models, namely Fitts Law, in Chapter 4. We describe its relevant application to reaching tasks in the field of Human Computer Interfacing (HCI) found in the literature, primarily with sighted users. We then show through analysis of experimental data how it is applied in comparison to the use of the hybrid control model proposed in Chapter 3.

Chapter 3 Hybrid System Model for Visual Control

In this chapter we discuss the various computer vision techniques that were employed for feature extraction and object identification. Different image feature sets were investigated for this work. The computer vision component of the system has been designed to be highly decoupled from the movement control component of the system. It allows for easy interchange of image feature extraction and matching techniques. The core requirement is that from the matched features, the moment based features can be calculated from the corresponding moment generating functions, $\mathbf{g}_1(f) = \begin{bmatrix} p_x \ p_y \end{bmatrix}^T$ and $\mathbf{g}_2(f) = f_z$ that will produce the image feature vector, $\mathbf{f} = \begin{bmatrix} p_x \ p_y \ f_z \end{bmatrix}^T$ that gives the terminal point for the current straight line trajectory of the reaching task.

We also introduce our hybrid control systems model based on a finite automata discrete event system (DES) that switches between the various controllers to perform the compensatory tracking of the axial components of the trajectory to the target. We also describe the tracking performance metrics proposed in the context of the DES model framework.

The computer vision component of the system resides on a personal computer and carries out the tasks of acquiring the sequence of images, analyzing each frame to detect the target object, and calculate the moment based features that determine the target's position within the image and distance from the camera.

A block diagram of the image analysis process for the vision system to perform object detection and control signal generation is given in Figure 3.1.

Image Frames → Noise Removal → Feature Extraction → Feature Matching → Object Localization → Control Input Generation
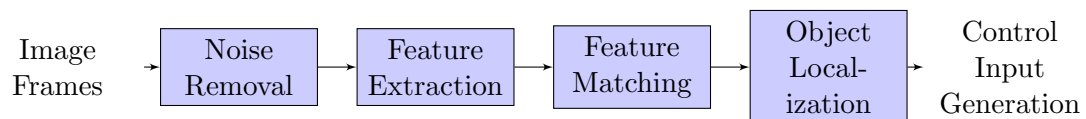
Figure 3.1: Image Processing Flow

3.1   Image Feature Extraction

Feature extraction is a important aspect of a visual control problem. When specifying the image features set, we must consider its use for both object identification, through feature matching, and object tracking/localization. For either purpose, various methods can be used employing point features, edge features, or region-based features; either separately or in combination.

The architecture for this system allows for a nearly complete decoupling of the image feature extraction and matching method from the rest of the system so that various algorithms can be used, interchangeably. The only necessary criteria is that, after matching is determined, a single three-element feature vector, f, can be calculated as the control input for servoing. The elements of the vector are the image plane coordinates of the centroid of the target object and a feature representative of the distance from the image to the target object along the optical axis. This will be discussed further in following sections of this chapter.

We examined various image feature sets in the development of this system. Since the focus of this work is to investigate the control for a reaching task through an assistive the device, our primary concern was not the development of any new feature extraction or matching techniques, but demonstrating the efficacy of the overall system. Three techniques used or examined were: colour segmentation, a combination of geometric shape features and colour segmentation, and Scale Invariant Feature Transform (SIFT).

3.2   Shape-Colour Feature Set Detection

The object detection approach is based on contour detection and geometrical shape classifications. The images acquired from the camera are passed through both a pyramid-up and pyramid-down operation for noise filtering prior to edge detection process. After which, Canny [82] edge detection is used to identify and index the resultant contours. Geometric shape identification is then carried out based on those contours. For example, a rectangle is characterized by four straight edges approximately at right angle to each other and four vertices. Due to inherent characteristics of image acquisition in an uncontrolled environment, for each object present in the scene more than one geometrical shape could be assigned. Therefore the list of shapes

initially identified is filtered to eliminate those that are collocated. The subsequent list of geometrical objects is then compared with a predetermined target object to identify which detected object has the highest priority of interest.

The level of importance of each feature was taken into consideration by assigning weights to features. For example if hue, saturation, and value are used as a subset of features, value has to be assigned a smaller weight as it varies significantly depending on the ambient lighting conditions. Using the feature values and the weights indicating the level of importance of each feature, a similarity measure can be calculated to determine which of the detected objects in the image has the highest priority of interest. The similarity measure, M, is given in (3.1).

$$M = \min_{k} \sum_{j=0}^{N} (w_j (fd_{kj} - ft_j)^2) \tag{3.1}$$

where fd_{kj} is the value of the jth feature of the kth detected object; ft_j is the value of the jth feature of the target object; w_j is the weight assigned to the jth feature; and N is the number of features considered. The feature values for the target object are static and set manually in advance.

The object found within the image with the lowest value of M is the best match (highest priority of interest) to the target object and thus becomes the object of interest. Having selected the object of interest, its features as well as location information with respect to the camera coordinates are extracted and the object is tracked in subsequent images.

The moment generating functions used to generate the feature vector, f, for the controller input are given by equation (3.2).

$$m_{ij} = \sum_{x} \sum_{y} x^i y^j I(x, y) \tag{3.2}$$

The centroid of the target object is then $\mathbf{g}_1(f) = [m_{10} \ m_{01}]^{\mathsf{T}}$. The term $I(x, y)$ is the intensity values of the within the bounds of the target object. The depth estimate feature is $\mathbf{g}_2(f) = \frac{m_{00}}{\alpha}$, where $0 \leq \alpha \leq 1$. In essence the value of α determines what portion of the mass of pixels with the bounds of target object fills the image frame so that the camera is sufficiently close to the target object for the reaching task to be considered complete. This scheme is quite limited in terms of

rotation of the target object pose and scale covariance, and requires calibration of the target object size to camera view at the task completion distance. However it does allow for target object detection in a moderately complex scene.

3.3 Feature Extraction and Tracking Using Scale-Space Theory

This visual feature tracker provides the control input to feature space controller. The task of this tracker is to detect the object of interest in the current frame and send the location and scale information to the controller, which is an image based look and move system. Given a particular object is of interest, the tracker detects the closest matching object in the current frame and keeps track of its location in subsequent frames. This is a challenging task. The appearance of the object of interest changes due to viewpoint and illumination. Moreover, occlusions are detrimental. Viewpoint change in general manifests locally in the form of an affine distortion [83]. Illumination changes are due to varying lighting itself and shadows. We disregard major affine distortions as the motion of the user is usually (by design) toward the object of interest. This prompts us to use a scale covariant and illumination invariant[1] feature detector.

Lack of scale covariance was one of the major drawbacks of our initial configuration [5] using the feature set scheme presented in the previous section. In that version of object detection, a pyramidal implementation of a polygon detector was used. As a result, only planar geometrical objects could be detected with that approach. Although a crude approximation of the scale was present in the level of pyramids used for contour extraction, that information was not effectively used. Therefore the system was not scale invariant. As a result the system had trouble in localizing the object of interest due to increased scale as the user moved toward the object. Color information complemented this drawback which itself needed illumination invariance. Consequently, the system was not robust. A scale and illumination invariant tracker which can handle general features is a desired solution.

Scale and illumination invariant feature detection techniques have improved tremendously during the last few years. Lindeberg, in his seminal work [84, 85],

---

1. From this point on we merely use the term invariant.

showed that features need to be detected within a scale-space framework for scale invariance. There are several well known methods for feature detection, computing local descriptors, and matching. Interested readers are referred to the recent surveys by Schmid *et al.* [86], Mikolajczyk and Schmid [87], Mikolajczyk *et al.* [88] and Moreels and Perona [89]. We adopt the so called difference-of-Gaussians for scale invariance as used by Lowe in his pioneering scale invariant feature transforms (SIFT) [90, 91]. In this section we outline the scale-space formulation and describe our adaptation of scale invariant object detection.

3.3.1   Scale-space Detection

In real images features exist in specific scales. Therefore, features need to be detected within the scale-space framework. The usual approach taken by feature detectors, especially in the case of distinctive features, is to find the local extremum in the scales-space [85]. Distinctive features such as SIFT [91] and *gradient location and orientation histogram* (GLOH) features [87] use the notion of scale-space to detect all the interest points (features) in an image. This approach has the advantage of being able to look at the image at different scales and a large number of features are detected. In a typical complex scene the number of features can range of in the thousands, so the approach is very computationally intensive. For example finding features in a typical 640×480 image would take an order of magnitude longer as compared to non-distinctive features [92] ($\sim$300 ms vs. $\sim$10 ms).

As indicated above, feature detection in real images needs to be done within a scale-space framework. It is even more critical when the need arises to match them to a different view. As in the case of this application the movement of the camera generates different views of the object. Features are seen only when viewed at a meaningful scale. Some features are seen at coarse scales and others are seen only at fine scales. For example, in an indoor scene, a door may be detected at a certain coarse scale. However, we need a finer-scale view to detect a name-plate on the door. Character or Braille data on the name-plate needs a finer scale. Even finer scales are needed to observe the texture on the door. Scale-space theory provides a general framework for dealing with images (multi-dimensional signals in general) at different scales in a well principled manner. However, the scale at which a particular feature is seen is unknown apriori. Therefore an automatic scale selection mechanism is used

to select the appropriate scale for each feature. This is done by finding local extrema with respect to scale in terms of Gaussian derivatives. Interestingly, the selected scale corresponds to the actual size of the structure (target object) in the image. This gives us the added advantage of being able to feed the scale of the object along with the location information to the feature space controller. Here we briefly outline the sale-space theory and automatic scale selection. A comprehensive description of this theory is given by Lindeberg [84, 85]. A general description is found in Bigun [93]. We follow Lindeberg's D-dimensional formulation. Adaption to 2-dimensional images is straight forward.

3.3.2 Scale-space Representation

Given any continuous signal $f : \mathbb{R}^D \to \mathbb{R}$, its linear scale-space representation $L : \mathbb{R}^D \times \mathbb{R}_+ \to \mathbb{R}$ is equivalent to the convolution of the function (image), $f(\mathbf{x})$, with Gaussian kernels, $h(\mathbf{x}; t)$, of various widths, t:

$$L(\mathbf{x}; t) = h(\mathbf{x}; t) * f(\mathbf{x}), \tag{3.3}$$

where $h : \mathbb{R}^D \to \mathbb{R}$ is given by,

$$h(\mathbf{x}; t) = \frac{1}{(2\pi t)^{D/2}} \exp\left(-\frac{x_1^2 + \cdots + x_D^2}{2t}\right) \tag{3.4}$$

and $\mathbf{x} = [x_1, \ldots, x_D]$. Scale-space derivatives are defined as

$$L_{x^\alpha}(\mathbf{x}; t) = (\partial_{x^\alpha} L)(\mathbf{x}; t) = \partial_{x^\alpha} (h(\mathbf{x}; t) * g(\mathbf{x})) \tag{3.5}$$

where $\alpha = [\alpha_1, \ldots, \alpha_D]$ and $\partial_{x^\alpha} L = L_{x_1^{\alpha_1} \ldots x_D^{\alpha_D}}$ constitute multi-index notation for the derivative operator ∂_{x^α}. Commutation implies

$$L_{x^\alpha}(\mathbf{x}; t) = (\partial_{x^\alpha} h(\cdot; t)) * g(\mathbf{x}). \tag{3.6}$$

So a set of Gaussian derivative kernels $\partial_{x^\alpha} h$ can be used to calculate scale-space derivatives efficiently. It is possible to obtain directional derivatives in the direction β by

$$\partial_\beta^n = (\cos \beta \partial_x + \sin \beta \partial_y)^n L. \tag{3.7}$$
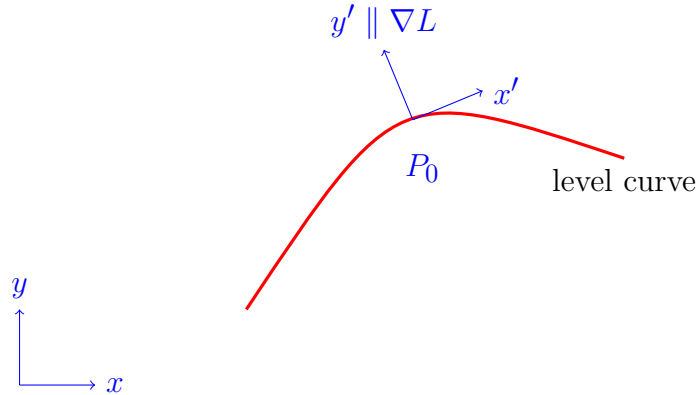
Figure 3.2: Local orthonormal coordinate frame

Equation (3.7) can be used to come up with a descriptor for a feature in an image. However, this will not make the descriptor rotational invariant. The way to define rotational invariant scale-space derivative based descriptors is by using a local coordinate system which is aligned with the local image feature itself. In particular, this defines a local orthonormal coordinate system (x', y') as shown in Fig. 3.2. The y'-axis is parallel to ∇L with $e_{y'} = (\cos\varphi + \sin\varphi)$ and $e_{x'} = (\sin\varphi - \cos\varphi)$, where $\varphi = \tan^{-1}(L_x/L_y)$. Now, quantities such as $L_{x'} = 0$, $L_{y'} = |\nabla L| = \sqrt{L_x^2 + L_y^2}$, $L_{y'}^2 L_{x'x'}$, etc. can be calculated. These quantities are invariant with respect to the rotation of image plane.

3.3.3 Automatic Scale Selection

Lindeberg [85] showed with examples that the local level of the scale crucially affects the performance of feature detectors. Therefore it is essential to perform the feature detection within a framework which automatically adapts the scale levels to the local image structure. A proven approach to perform this adaptation is by detecting local extrema over scales of normalized differential entities giving rise to a characteristic scale. Normalization is required due to the property that the scale-space spatial derivatives decrease in magnitude with increasing scale. The *γ-normalized derivative operator* defined by

$$\partial_{\xi, \gamma-norm} = t^{\gamma/2} \partial_x \qquad (3.8)$$

is used for this purpose. With this normalization, maxima over scales of normalized derivatives reflect the scales over which spatial variations take place in the signal.

Therefore, the scale selection principle [85, 94] is, in the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data. Automatic scale selection makes an interest point scale invariant. However, affine invariance is not guaranteed and feature repeatability decreases with large viewpoint changes [83].

3.3.4 SIFT Features

Automatic scale selection by scale-space extrema detection is what drives the interest point identification known as Scale Invariant Feature Transform (SIFT) scheme developed by Lowe [91] . These 'good' features [95] are called key points. Once the key points are identified, the accurate location and scale is determined. The next step is the orientation assignment. Finally, a local image gradient based descriptor, which is a 128-dimensional vector, is calculated. SIFT does the scale-space extrema detection by approximating equation (3.6) using the difference between $L(\mathbf{x}; k^2 t)$ and $L(\mathbf{x}; t)$, where k is a constant multiplicative factor. This quantity is called the difference of Gaussian and is expressed by,

$$
\begin{aligned}
D(\mathbf{x}; k\sigma) &= \left(h(\mathbf{x}; k^2\sigma^2) - h(\mathbf{x}; \sigma^2) \right) * f(\mathbf{x}) \\
&= L(\mathbf{x}, k^2\sigma^2) - L(\mathbf{x}; \sigma^2),
\end{aligned}
\tag{3.9}
$$

where $\sigma^2 = t$. Lowe [91] shows that

$$
h(\mathbf{x}; k^2\sigma^2) - h(\mathbf{x}; \sigma^2) \approx (k-1)\sigma^2 \nabla^2 h
\tag{3.10}
$$

which means this variant inherently incorporates the $t = \sigma^2$ normalization required for scale-invariant Laplacian. A constant $k - 1$ is the same for all scales and therefore does not affect the extrema detection. The Difference of Gaussians are calculated by computing a stack of Gaussians and then doing image subtraction. A resampling process is used when moving from one octave to the next, which reduces computations. Interested readers are referred to Lowe's paper for details [91]. Local extrema are detected within this stack of difference of Gaussian images. This is done by comparing each sample point $D(\mathbf{x}, k\sigma)$ against its 26 neighbors. Such a local extrema is an interest point. However, the accurate key point location is further refined by fitting a

3-D quadratic function to the nearest sample points and interpolating. The resulting key point location is therefore sub-pixel accurate [96]. Key points which correspond to low contrast points and those on edges are not useful. These can be rejected by computing the autocorrelation matrix [97] of $D(\mathbf{X}; k\sigma)$ (call this A_D) and ensuring the trace-to-determinant ratio,

$$\frac{\mathrm{T}(A_D)}{|A_D|} < \frac{(r+1)^2}{r} \tag{3.11}$$

where r is the threshold for the eigenvalue ratio of A_D. Lowe [91] choose to use $r = 10$. What remains to be assigned is the orientation. Orientation is assigned by finding the parabolic-interpolated dominant gradient direction calculated using a 32-bin gradient location and orientation histogram.

Scale-space extrema detection automatically assigns a scale to every key point. We outlined above how the location and orientation of a key point is found. The next step is to compute a distinctive descriptor[2] for the local image region which can also be matched in a view-point invariant manner. First the image gradient magnitude and orientations are samples around the key point location. Orientation invariance is achieved by transforming the gradients on to the local coordinate frame oriented along the dominant gradient direction calculated in the orientation assignment phase. In addition, the gradient magnitudes are Gaussian weighted. The descriptor is a summary of the gradient magnitude orientations: a 3-D histogram of gradient location and orientations. In practice, a 16×16 region, 4×4 subregions, and 8 orientation bins are used [91]. This gives rise to the infamous $4 \times 4 \times 8 = 128$-dimensional vector descriptor. This vector is thresholded and normalized to unit length to suppress the effects of illumination changes, resulting in partial illumination invariance.

3.3.5 Implementation

We described how SIFT features are a good candidate as features for the visual tracker in complex natural scenes. However, the features need to be strategically used in order to be useful for our system. First we assume that we know what the object of interest is. In other words, features corresponding to the object of interest

---

2. See Mikolajczyk 2005 for a comparison of local region descriptors [87]. They found that gradient location and orientation histogram used in SIFT to be a good descriptor.

on the initial video frame are assumed to be known. Recent work on video google [98, 99] provides an interesting method of automating this task. To that end, a bag of features for objects are stored on disk. Once the target object is indicated by user, potential features that correspond to this object can be selected. Once the presence of these features in a geometrically coherent manner on the current image is verified, object tracking can take place. Since our implementation requires manual selection, we assume the features corresponding to the object of interest are known. We locate them on subsequent images only within a subspace of the scale-space. This subspace is selected using the knowledge of the scale of the object on the current image and the approximate location. Approximate location selection is valid only if the motion of the user's arm is small. If the features corresponding to the object is not found within this subspace, we search a larger space containing the whole image plane. When there is more than one feature corresponding to the object of interest, we treat the centroid of the set of features as the location. The scale of the coarsest feature is used as the nominal scale of the object of interest.

3.4 Colour Segmentation based Feature Set Detection

A third, fairly simplistic feature set was employed for the purpose of maximizing the visual tracker frame rate. Since motion cues are issued on a per frame basis the purpose was to minimize any response rate limitations on the resulting arm movement by the user. Target object detection was done by simple colour segmentation and the scene was tightly constrained to a set of uniformly coloured spheres.

The moment generating functions used to generate the feature vector, f, for the controller input are similar to that previously given in equation (3.2). The centroid of the target object is then $\mathbf{g}_1(f) = [m_{10}\ m_{01}]^\mathsf{T}$. The term $I(x, y)$ is the intensity values of the target colour within the bounds of the target object. The depth estimate feature is $\mathbf{g}_2(f) = \frac{m_{00}}{\alpha}$, where $0 \leq \alpha \leq 1$. Similar to the feature set describe in Section 3.2, the mass of pixels with the bounds of target object that fills the image frame directly relates to a sufficiently close distance from the target object to consider the reaching task as complete. It also still requires calibration of the target object size to camera view at the task completion distance. However, since the objects are spheres there is

no longer the same limitations with rotation and scale invariance allowing for greater freedom in experimental trials.

3.5 Camera Geometry and Scale

When we attempt to use the scale parameter of a feature to be representative of the size of the object, we treat the camera as a measurement device. We assume that the rays pass through the optical center of the camera. Such a camera is a central projection device. In this section we outline the central projective geometry, and show how the scale of feature (or an object) varies with the camera configuration. We follow Hartley and Zisserman [100] in this section.

3.5.1 Camera Models

A camera maps 3-space world points to 2-space points in the image. In central projection, the world point, the camera center, and the image of the world point are collinear. Figure 3.3 shows the central projection (pinhole) camera model. Point $\mathbf{C}$ is the camera center (center of projection or optical center). The camera center coincides with the world coordinates origin[3]. The optical axis is the line through this point that is perpendicular to the image plane. The point at which the optical axis meets the image plane, $\mathbf{P}$, is the principle point. Focal length f is the distance from the camera center to the principle point. This camera maps a world point $\mathbf{X} = [X, Y, Z]^\mathsf{T}$ to the point $\mathbf{x} = [fX/Z, fY/Z]^\mathsf{T}$ in the image plane, if we assume that $\mathbf{P}$ is the image origin. Now we can express this mapping in homogeneous coordinates as

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.12}$$

---

3. This assumption does not harm the generality.
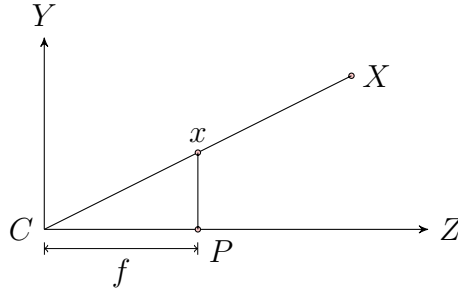
Figure 3.3: Camera geometry – $\mathbf{C}$ is the camera center and $\mathbf{C}Z$ is the optical axis. $\mathbf{x}$ is the image of the world point $\mathbf{X}$. f is the focal length and $\mathbf{x}$ has image coordinates $(f\frac{X}{Z}, f\frac{Y}{Z})$ if the image center is $\mathbf{P}$ and image coordinate axes are X and Y.

The 3×4 matrix $[\text{diag}(f, f, 1)|\mathbf{0}]$ is called the camera matrix [100]. If $\mathbf{P}$ has image coordinates $\left[p_x, p_y\right]^{\mathsf{T}}$, then the mapping is

$$\mathbf{X} \mapsto \begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.13}$$

Rearranging,

$$\mathbf{X} \mapsto \begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} [\mathtt{I}|\mathbf{0}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.14}$$

The matrix

$$\mathtt{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \tag{3.15}$$

is called the camera calibration matrix. Note that we have assumed equal focal lengths (f in pixels) and no skew parameter for simplicity.

When we consider, for example, the x-values of the image point $\mathbf{x}$, we notice that the graph between x and Z is a rectangular hyperbola in the first quadrant. If we consider a world sphere of diameter D, the diameter of its image (a circle) d is related to the distance Z by

$$d = \frac{fD}{Z}. \tag{3.16}$$

In this equation fD is a constant for a given sphere and a camera. One verifies this by noting that the presence of the principal point in (3.14) does not affect the diameter. In other words, diameter of the sphere seen in the image varies reciprocally with the distance form the optical center. If we can approximate the diameter with a scale parameter, we can make approximations to the distance once a scale-distance calibration is done.

3.6   Hybrid Control System Model Using Image Features

From the material presented in the previous chapters it should be evident that a conventional continuous time-invariant model cannot appropriately capture the behaviour of this system. Without an accurate dynamic model of the plant during the reaching nor the ability to drive the plant through a desired trajectory we must model the system in a manner that captures what the user is "attempting" to do in terms of the compensatory tracking task. This can be delineated as a set of events occurring throughout a basic set of spatial tracking operation states: resolving horizontal alignment error, resolving vertical alignment error, and resolving depth alignment error. Thus a problem can be modeled using hybrid system approach to describe the tracking operation tasks as discrete automaton and actuating the appropriate continuous-time control law for a particular tracking task.

We have chosen to use a hybrid control scheme to model the system's behaviour and we provide a brief description of it in this section so that the reader can form a clear picture of the overall system architecture. The formalism and notation used here is adopted from the class of supervisory control based Discrete Event Systems (DES) models proposed by Stiver and Antsaklis in several papers [101, 102, 103, 104, 105, 106], with an expanded treatment provided Koutsoukos *et al.* [107] a few years later. The reader is referred there for further details of the formalism.

A block diagram of the hybrid control system model components is provided in Figure 3.4. The system consists of three components: a continuous-time plant, a DES-controller, and the interface between the two which converts between the requisite continuous-time signals and discrete symbols for each corresponding input and output.
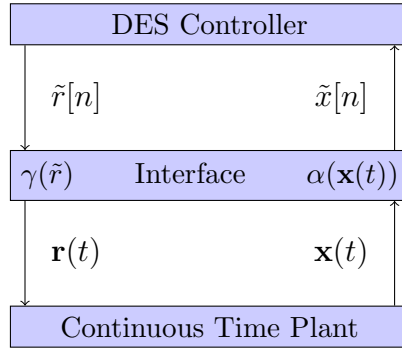
Figure 3.4: Block diagram of the discrete-to-continuous mapping functions (interface) between continuous-time plant (CT-plant) and discrete event system controller (DES-controller)

The continuous-time plant and the Interface taken together as one unit form the DES-plant paired to the DES-controller. First we will describe each of the three components of the model formalism and their interaction in general terms, then provide details specific to the control problem we are addressing in this work.

A) Continuous-Time Plant (CT-plant): is generally a nonlinear, time-invariant system expressed as

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{r}) \tag{3.17}$$

$$\mathbf{y} = g(\mathbf{x}) \tag{3.18}$$

where $\mathbf{x}\epsilon\mathbb{R}^n$, $\mathbf{r}\epsilon\mathbb{R}^m$, and $\mathbf{y}\epsilon\mathbb{R}^p$ are the state, input, and output vectors, respectively. The input, $\mathbf{r}(t)$, is a piecewise continuous signal issued by the interface and based on the DES-controller output symbols. Unless otherwise explicitly stated elsewhere, we take the output function to be, $g(\mathbf{x}) = \mathbf{x}$.

B) Interface: the interface between the DES-controller and the CT-plant consists of two memoryless mapping functions that preform the continuous signal to discrete symbol conversion and vise versa. The actuator function, $\gamma : \tilde{R} \rightarrow \mathbb{R}^m$, given by

$$\mathbf{r}(t) = \gamma(\tilde{r}) \tag{3.19}$$

maps the sequence of DES-controller output symbols to a piecewise-continuous input

signal for the CT-plant. Similarly, the generator function, $\alpha : \mathbb{R}^n \to \tilde{X}$, given by

$$\tilde{x} = \alpha(\mathbf{x}(t)) \tag{3.20}$$

maps the state-space of the CT-plant to the set of plant event symbols for the DES-plant.

  C) DES-controller: the DES controller is specified by a quintuple, $\{\tilde{S}, \tilde{X}, \tilde{R}, \delta, \phi\}$, where $\tilde{S}$ is the set of DES-controller states, $\tilde{X}$ is the set of DES-plant event symbols, and $\tilde{R}$ is the set of DES-controller output symbols. The DES-controller state transition function, $\delta : \tilde{S} \times \tilde{X} \to \tilde{S}$, and the DES-controller output function, $\phi : \tilde{S} \to \tilde{R}$, describe the behaviour of the DES-controller and are given by the following equations,

$$\tilde{s}[n] = \delta(\tilde{s}[n-1], \tilde{x}[n]) \tag{3.21}$$

$$\tilde{r}[n] = \phi(\tilde{s}[n]) \tag{3.22}$$

  Where $\tilde{s}[n] \, \epsilon \, \tilde{S}$, $\tilde{x}[n] \, \epsilon \, \tilde{X}$, $\tilde{r}[n] \, \epsilon \, \tilde{R}$, and n indexes the order of symbols occurring in the sequence of events. The symbols in $\tilde{R}$ represent the DES-controller action to be taken and symbols in $\tilde{X}$ correspond to events occurring in the CT-plant state-space. DES-plant event symbols are generated when the state of the CT-plant crosses from one region of its state-space into another region. The regions are partitioned by hypersurfaces that are specified as the boundary, which when crossed, triggers the occurrence of the corresponding DES-plant event. The set of DES-plant events recognized by the generator function are defined by the set of hypersurface functions, $\{h_i : \mathbb{R}^n \to \mathbb{R}, i \epsilon I\}$, that must satisfy the following conditions

$$\nabla_x h_i(\xi) \neq 0, \qquad \forall \xi \epsilon \aleph(h_i) \tag{3.23}$$

that stipulate the null-space of the functions, $\aleph(h_i) = \{\xi \epsilon \mathbb{R}^n : h_i(\xi) = 0\}$, forms an $n - 1$ dimensional smooth hypersurface separating the state-space. Koutsoukos *et al.* define that if the hypersurface derivative is nonzero at the crossing, the conditions can be simplified so that generation of a CT-plant event can be expressed as

$$h_i(\mathbf{x}(t)) = 0, \qquad \frac{\partial}{\partial t} h_i(\mathbf{x}(t)) \neq 0 \tag{3.24}$$

The following subsections will describe the various components of our hybrid control

system using the formalism provided above.

3.6.1 The 1-D DES-Controller Model

Since the user follows an egocentric reference frame, terms such as left, right, up, down, and forward with be used to reference directional motion. We start by constructing a model for a one dimensional compensatory tracking problem in the context of horizontal movement in the image. The state-space is simply a line segment that is partitioned by two hypersurfaces (points) bounding the approximate target region. We define the line to be I_x in length (width of the image) and specify the precise target registration point at $\frac{I_x}{2}$. The position of the hypersurface boundaries on the line segment are specified by the parameter, l, and given by, $x_{p1} = \frac{I_x}{2}(1 - l)$ and $x_{p2} = \frac{I_x}{2}(1 + l)$. An illustration is shown in Figure 3.5, with $0 < l \leq 1$. With this we conform to the need for allowing for an approximate targeting region discussed previously in Section 2.4. As $l \to 0$ the On-Target region reduces to a single point, and conversely as $l \to 1$ the On-Targetregion becomes the entire image width.

$$x = x_{p1} \qquad \frac{I_x}{2} \qquad x = x_{p2}$$

$$\tilde{x}_1 \qquad \tilde{x}_3$$

$$\tilde{x}_2 \qquad \tilde{x}_4$$
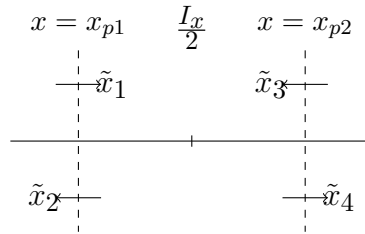
Figure 3.5: State space diagram for 1D *approximate* targeting model

Two hypersurface functions are defined at each boundary point to support the events triggers that correspond to crossings in a specific direction. The four functions, $h_1(x)$ through $h_4(x)$ are given by

$$h_1(x) \;=\; x - \frac{I_x}{2}(1 - l) \tag{3.25}$$

$$h_2(x) \;=\; \frac{I_x}{2}(1 - l) - x \tag{3.26}$$

$$h_3(x) \;=\; \frac{I_y}{2}(1 + l) - x \tag{3.27}$$

$$h_4(x) \;=\; x - \frac{I_y}{2}(1 + l) \tag{3.28}$$

The set of DES-plant events, $\tilde{X}$, that are trigger upon crossings are defined as

$\tilde{x}_1$ `On-Target` occurs after crossing $h_1(x)$ by a movement Left;

$\tilde{x}_2$ `Off-Target` occurs after crossing point $h_2(x)$ by a movement Right;

$\tilde{x}_3$ `On-Target` occurs after crossing point $h_3(x)$ by a movement Right;

$\tilde{x}_4$ `Off-Target` occurs at crossing point $h_4(x)$ by a movement Left;

The set of DES-controller states for this model consist of the following symbols (descriptions): $\tilde{s}_1$ (`On-Target`), $\tilde{s}_2$ (`Off-Target`-Right), and $\tilde{s}_3$ (`Off-Target`-Left). The DES-controller output function, ϕ, given by equation (3.22) will generate the corresponding output symbols (descriptions): $\tilde{r}_1$ (`Stop`), $\tilde{r}_2$ (`Move-Left`), and $\tilde{r}_3$ (`Move-Right`). The DES-controller state transitions defined by the function, δ, are illustrated in Figure 3.6. Since this is a compensatory tracking problem in which the target is stationary (object on a shelf) and the end-effector is moving into alignment, we define the controller output symbols to reference the direction of movement of end-effector. Thus when the centroid of the target object within the image frame is left of center the end-effector (camera) must move to the left to align center of the frame. The transitions from $\tilde{s}_2$ and $\tilde{s}_3$ to $\tilde{s}_1$ depict DES-controller state transitions
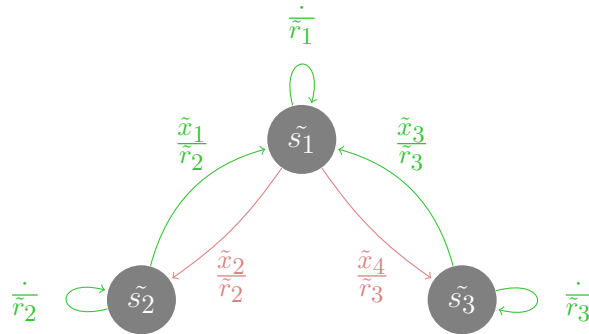
Figure 3.6: DES-controller state transition diagram for 1D *approximate* targeting l-model

from `Off-Target` regions to the `On-Target` region. The transition of $\tilde{s}_1$ looping back upon itself indicates that the DES-controller maintains an `On-Target` state regardless of slight position jitter of the CT-plant as the user attempts to hold their arm steady within the target region. This is considered part of the ideally-behaved

set of state transitions as it is an expected and acceptably stable result of the plant dynamics.

In close proximity to either hypersurface, position jitter and susceptibility to other disturbances can cause a problem similar to precise alignment to a single point. If the user is tracking at a sufficiently slow rate and is able to respond to a controller output of $\tilde{r}_1$ (Stop) just within the bounds of the target region, there is a reasonable possibility that the user could drift back over the target region boundary, generating an Off-Target based event. This results in an inefficient, possibly unstable trajectory through the state-space. The problem can be amplified by successive response, and drift back and forth across the boundary producing an oscillating sequence of DES-plant events and DES-Controller symbols in result.

Unstable behaviours such as overshoot of the target region cannot be represented by direct state transition as in equation (3.21). Overshoot with a well-behaved user can occur if l is sufficiently small and/or end-effector velocity is sufficiently large so that momentum carries it across the opposite hypersurface before the DES-plant can respond to the $\tilde{r}_1$ (Stop) control symbol. This result would be shown after either of the sequences of state transitions: $\frac{(\tilde{s}_2:\tilde{x}_1)}{\tilde{r}_2} \rightarrow \frac{(\tilde{s}_1:\tilde{x}_4)}{\tilde{r}_1}$ or $\frac{(\tilde{s}_3:\tilde{x}_1)}{\tilde{r}_3} \rightarrow \frac{(\tilde{s}_1:\tilde{x}_2)}{\tilde{2}_2}$. The same could occur for an ill-behaved user if the dynamics were sufficiently slow so that the DES-plant could respond to the Stop command, but did not out of choice[4]. Regardless of which scenario was the cause of an overshoot, DES-plant events and DES-controller states are queued and processed in sequence.

Since the direction of the state-space trajectory at a hypersurface crossing determines the DES-plant event symbol, certain direct state transitions are impossible. In this case, $(\tilde{s}_2, \tilde{x}_2)$ and $(\tilde{s}_3, \tilde{x}_4)$ as they represent the trajectory of approaching one side of a hypersurface boundary yet instantaneously crossing in the opposite direction from the other side.

3.6.2 The 2-D DES-Controller Model

Extending the description of the state-space to the 2-D image frame and using the parameter, l, in a similar fashion to equation (2.4) we can define the four pairs of coincident state-space partitioning functions that create an approximate target

---

4. We do not attempt to chart the possible reasons when a user chooses to not respond to a control signal, only that it is a form of disturbance that can occur

region. We define $\mathbf{x} = [x_1 \ x_2]^\mathsf{T}$ as the position state vector and using I_x and I_y as the image dimensions in pixels, the hypersurface boundaries are given by equations (3.29) through (3.36).

$$h_1(\mathbf{x}) = x_1 - \frac{I_x}{2}(1-l) \tag{3.29}$$

$$h_2(\mathbf{x}) = \frac{I_x}{2}(1-l) - x_1 \tag{3.30}$$

$$h_3(\mathbf{x}) = \frac{I_x}{2}(1+l) - x_1 \tag{3.31}$$

$$h_4(\mathbf{x}) = x_1 - \frac{I_x}{2}(1+l) \tag{3.32}$$

$$h_5(\mathbf{x}) = x_2 - \frac{I_y}{2}(1-l) \tag{3.33}$$

$$h_6(\mathbf{x}) = \frac{I_y}{2}(1-l) - x_2 \tag{3.34}$$

$$h_7(\mathbf{x}) = \frac{I_y}{2}(1+l) - x_2 \tag{3.35}$$

$$h_8(\mathbf{x}) = x_2 - \frac{I_y}{2}(1+l) \tag{3.36}$$

A diagram of the image state-space partitioned by the eight hypersurfaces (lines) is provided in Figure 3.7. It shows which DES-plant events are trigger as the target point $\mathbf{p}_{xy} = [x_1(t) \ x_2(t)]^T$ moves through the image. Individual events are defined for unidirectional crossing of hypersurface segments bounded by either the image boundaries or the intersections with the other hypersurfaces. The descriptions of the DES-plant events and related camera motion are given in Table 3.1.

The DES-controller state symbols, $\tilde{S}$, the controller state, and descriptions of the target's relative position are given in table 3.2. There are ten DES-Controller states, with $\tilde{s}_1$ through $\tilde{s}_9$ representing the various states of tracking towards the target point $\mathbf{p}_{xy}$ and $\tilde{s}_{10}$ representing the absence of a target. The mapping for the DES-controller output function, $\phi : \tilde{S} \to \tilde{R}$, is given in Table 3.3. The second and third rows of the table indicate that multiple DES-controller states produce the same output symbol. This indicates how we prioritize horizontal target alignment over vertical alignment in the two dimensional case.

A state transition diagram for the DES-controller is given in Figure 3.8. Both the DES-plant event symbol and corresponding DES-controller output symbol are provided for each transition. Each transition arc is colour coded to indicate which

Table 3.1: DES-plant event $(\tilde{X})$ symbol table for two dimensional tracking with the l-model

Symbol	DES-Plant Event Symbols	Hypersurface	Camera Movement
$\tilde{x}_1$, $\tilde{x}_5$, $\tilde{x}_9$	On-Target in horizontal $(x_1(t))$	$h_1(\mathbf{x}) > 0$	left
$\tilde{x}_2$, $\tilde{x}_6$, $\tilde{x}_{10}$	Off-Target in horizontal $(x_1(t))$	$h_2(\mathbf{x}) < 0$	right
$\tilde{x}_3$, $\tilde{x}_7$, $\tilde{x}_{11}$	On-Target in horizontal $(x_1(t))$	$h_3(\mathbf{x}) > 0$	right
$\tilde{x}_4$, $\tilde{x}_8$, $\tilde{x}_{12}$	Off-Target in horizontal $(x_1(t))$	$h_4(\mathbf{x}) < 0$	left
$\tilde{x}_{13}$, $\tilde{x}_{17}$, $\tilde{x}_{21}$	On-Target in vertical $(x_2(t))$	$h_5(\mathbf{x}) > 0$	down
$\tilde{x}_{14}$, $\tilde{x}_{18}$, $\tilde{x}_{22}$	Off-Target in vertical $(x_2(t))$	$h_6(\mathbf{x}) < 0$	up
$\tilde{x}_{15}$, $\tilde{x}_{19}$, $\tilde{x}_{23}$	On-Target in vertical $(x_2(t))$	$h_7(\mathbf{x}) > 0$	up
$\tilde{x}_{16}$, $\tilde{x}_{20}$, $\tilde{x}_{24}$	Off-Target in vertical $(x_2(t))$	$h_8(\mathbf{x}) < 0$	down

of the three behavioural classes it corresponds to. Green transitions label ideally-behaved state transitions which follow the horizontally prioritized tracking trajectory to the target region. Blue arcs label transitions of well-behaved tracking that break from the horizontal priority, but still minimize the overall planar distance to the target. The transitions that are labeled in red specify the ill-behaved state transitions where the motion of the hand/camera actually increases the planar distance to the target. While it is likely the non-ideal transitions are likely due to minor unintentional horizontal and/or vertical position drift of the camera, it is also possible that they could be attributed to intentional movements that are in contradiction to the DES-controller output commands. The Halt state, $\tilde{s}_{10}$, and the loop back transition on $\tilde{s}_1$ were omitted from the diagram for to avoid visual clutter. All states can transition to $\tilde{s}_{10}$ through at any time, with or without a triggered plant, $\dot{\overline{\tilde{r}_6}}$, if the target object is lost from the image.

Table 3.2: DES-controller state ($\tilde{S}$) symbol table for two dimensional tracking with the l-model

Symbol	State Description	Relative Target Position
$\tilde{s}_1$	On-Target	p_{xy} in target region
$\tilde{s}_2$	Off-Target-Right	p_{xy} Left of target region
$\tilde{s}_3$	Off-Target-Left	p_{xy} Right of target region
$\tilde{s}_4$	Off-Target-Above	p_{xy} Below of target region
$\tilde{s}_5$	Off-Target-Right-Above	p_{xy} Left-Below of target region
$\tilde{s}_6$	Off-Target-Left-Above	p_{xy} Right-Below of target region
$\tilde{s}_7$	Off-Target-Below	p_{xy} Above of target region
$\tilde{s}_8$	Off-Target-Right-Below	p_{xy} Left-Above of target region
$\tilde{s}_9$	Off-Target-Left-Below	p_{xy} Right-Above of target region
$\tilde{s}_{10}$	NO-Target	

Table 3.3: Table of DES-controller output ($\tilde{R}$) symbols, given by equation (3.22), for two dimensional tracking with the l-model

Symbol	DES-controller Output Symbols
$\phi(\tilde{s}_1) = \tilde{r}_1$	Stop: successful reach
$\{\phi(\tilde{s}_2), \phi(\tilde{s}_5), \phi(\tilde{s}_8)\} = \tilde{r}_2$	Move-Left
$\{\phi(\tilde{s}_3), \phi(\tilde{s}_6), \phi(\tilde{s}_9)\} = \tilde{r}_3$	Move-Right
$\phi(\tilde{s}_4) = \tilde{r}_4$	Move-Down
$\phi(\tilde{s}_7) = \tilde{r}_5$	Move-Up
$\phi(\tilde{s}_{10}) = \tilde{r}_6$	Halt: no target

Figure 3.7: Image state-space diagram for two dimensional *approximate* targeting using the *l*-model
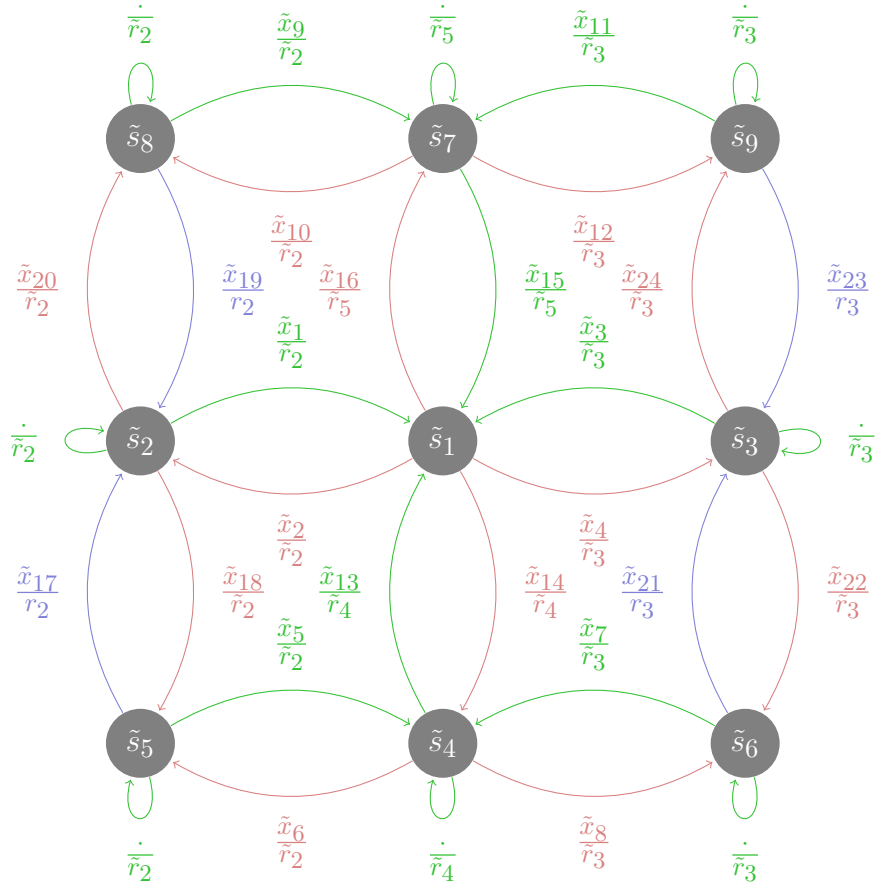
Figure 3.8: DES-controller state transition diagram for two dimensional l-model. Green indicates-ideally behaved transitions, blue indicates well-behaved transitions, and red indicates ill-behaved transitions

3.6.3 The 1-D DES-Controller Model with Hysteresis

To alleviate the potential drift instability an alternate method of partitioning the state-space can be used. We call this the l_2-model. It involves additional hypersurfaces and redefinition of the generator function, $\alpha(\mathbf{x}(t))$ (equation (3.20)), so that the sequence of DES-plant event symbols can represent two types of events based on the direction in which a hypersurface is crossed. The new definition is given in equation (3.37),

$$\tilde{x}[n] = \begin{cases} \alpha(\mathbf{x}(\tau_e[n])) & \text{non-silent event} \\ \varepsilon_i & \text{silent event} \end{cases} \tag{3.37}$$

where $\tau_e[n]$ is the time stamp of the n-th event in the trajectory traversal sequence. The symbol ε denotes a null symbol (silent event). With this definition the generator function triggers non-silent events when a hypersurface is crossed in a one direction, but a silent event when it is crossed in the opposite direction. Silent events have no effect on DES-controller state transitions, so $\tilde{x}[n] = \alpha(\mathbf{x}(\tau_e[n-1]))$. To provide an example we partition the one-dimensional state-space with a set of four hypersurfaces given in equations (3.38) through (3.41), below. The placement of the four hypersurfaces are specified by the two parameters l_1 and l_2. Similar to Section 3.6.1, $0 < l_1, l_2 \leq 1$.

$$h_1(x) = x - \frac{I_x}{2}(1 - l_1) \tag{3.38}$$

$$h_2(x) = \frac{I_x}{2}(1 - l_1 + l_1 l_2) - x \tag{3.39}$$

$$h_3(x) = \frac{I_x}{2}(1 + l_1) - x \tag{3.40}$$

$$h_4(x) = x - \frac{I_x}{2}(1 + l_1 + l_1 l_2) \tag{3.41}$$

They are illustrated in Figure 3.9. The two `On-Target` events are triggered at the crossing of either $h_1(x)$ or $h_3(x)$ towards the center point, but only a silent event is generated if the trajectory were to drift back over either hypersurface, so the DES-controller state remains as $\tilde{s}_1$. This imposes a hysteresis on the target acquisition trajectory through the state-space. The hysteresis effect also applies with overshoot of the target region. If tracking `On-Target` from a crossing of h_1 and a subsequent crossing of h_3 occurs before the plant responds to the `Stop` command, the `On-Target` state would still be valid as long as h_4 was not crossed.

With non-silent events taking precedence over simultaneous silent events and letting $l_2 \rightarrow 0$, the hysteresis partitioning reduces to the basic one-dimensional l-model previously described in Section 3.6.1. When $l_1 \rightarrow 0$, the set of hypersurfaces align to the center of the frame so that the generator function triggers `On-Target` events as if the plant were an ideal robot capable of precise target registration.
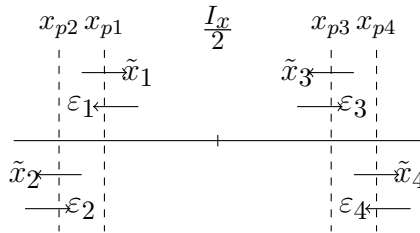
Figure 3.9: State space diagram for the one dimensional l_2 model for target region hysteresis.

The DES-controller state transition function, δ and DES-controller output function, ϕ, for the hysteresis partitioning are illustrated in Figure 3.10. As in the l-model case, the DES-controller states are still represented by the symbols: `On-Target` $(\tilde{s}_1)$, `Off-Target`-Left $(\tilde{s}_3)$, and `Off-Target`-Right $(\tilde{s}_2)$. But there is the addition of loop-back transitions based on the silent event crossings.
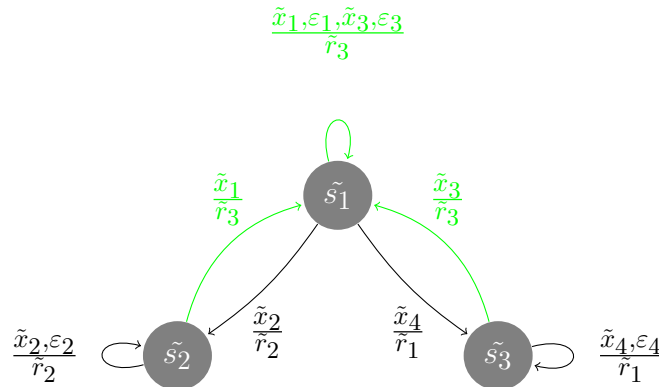
Figure 3.10: DES-controller state transition diagram for the one dimensional l_2-model for target region hysteresis.

3.6.4 The 2-D DES-Controller Model with Hysteresis

We can extend the DES to an l_2-model for two dimensions by specifying the following functions (equations (3.42) through (3.49)) for partitioning the state-space. In a similar manner to the one dimensional case, the hypersurface position parameters are $0 \leq l_1, l_2 \leq 1$. An illustration of the state-space with both the non-silent and silent DES-plant event symbols labeling the directional crossing is given in Figure 3.11.

$$h_1(\mathbf{x}) = x_1 - \frac{I_x}{2}(1 - l_1) \tag{3.42}$$

$$h_2(\mathbf{x}) = \frac{I_x}{2}(1 - l_1 + l_1 l_2) - x_1 \tag{3.43}$$

$$h_3(\mathbf{x}) = \frac{I_x}{2}(1 + l_1) - x_1 \tag{3.44}$$

$$h_4(\mathbf{x}) = x_1 - \frac{I_x}{2}(1 + l_1 + l_1 l_2) \tag{3.45}$$

$$h_5(\mathbf{x}) = x_2 - \frac{I_y}{2}(1 - l_1) \tag{3.46}$$

$$h_6(\mathbf{x}) = \frac{I_y}{2}(1 - l_1 + l_1 l_2) - x_2 \tag{3.47}$$

$$h_7(\mathbf{x}) = \frac{I_y}{2}(1 + l_1) - x_2 \tag{3.48}$$

$$h_8(\mathbf{x}) = x_2 - \frac{I_y}{2}(1 + l_1 + l_1 l_2) \tag{3.49}$$

The partitioning of the state-space forms two rectangular bounding regions. The portion of the state-space within the inner boundary forms the definite approximate `On-Target` region and the portion of the state-space outside the outer boundary forms the definite `Off-Target` region. The portion of the state-space between the two boundaries is the hysteresis region which is considered `On-Target` if the tracking trajectory had previously crossing into the `On-Target` region through the proper sequence of DES-controller state transitions. The set of DES-controller state symbols, $\tilde{S}$, and the DES-controller output function, $\phi(\tilde{s})$, are the same as those defined in Tables 3.2 and 3.3, respectively. The DES-controller state transitions are illustrated in Figure 3.12 with green paths indicating ideally-behaved transitions, blue paths indicating well-behaved transitions, and red paths indicating ill-behaved transitions. Ideally-behaved transitions follow trajectories that are moving towards resolv-

Table 3.4: DES-plant event $(\tilde{X})$ symbol table for one dimensional tracking with the l_2-model

Symbol	DES-Plant Event Symbols	Hypersurface	Camera Movement
$\tilde{x}_1, \tilde{x}_5, \tilde{x}_9$	On-Target in horizontal $(x_1(t))$	$h_1(\mathbf{x}) > 0$	left
$\varepsilon_1, \varepsilon_5, \varepsilon_9$	remain On-Target in $x_1(t)$	$h_1(\mathbf{x}) < 0$	right
$\tilde{x}_2, \tilde{x}_6, \tilde{x}_{10}$	Off-Target in horizontal $(x_1(t))$	$h_2(\mathbf{x}) < 0$	right
$\varepsilon_2, \varepsilon_6, \varepsilon_{10}$	remain Off-Target in $x_1(t)$	$h_2(\mathbf{x}) < 0$	left
$\tilde{x}_3, \tilde{x}_7, \tilde{x}_{11}$	On-Target in horizontal $(x_1(t))$	$h_3(\mathbf{x}) > 0$	right
$\varepsilon_3, \varepsilon_7, \varepsilon_{11}$	remain On-Target in $x_1(t)$	$h_3(\mathbf{x}) < 0$	left
$\tilde{x}_4, \tilde{x}_8, \tilde{x}_{12}$	Off-Target in horizontal $(x_1(t))$	$h_4(\mathbf{x}) < 0$	left
$\varepsilon_4, \varepsilon_8, \varepsilon_{12}$	remain Off-Target in $x_1(t)$	$h_4(\mathbf{x}) < 0$	right
$\tilde{x}_{13}, \tilde{x}_{17}, \tilde{x}_{21}$	On-Target in vertical $(x_2(t))$	$h_5(\mathbf{x}) > 0$	down
$\varepsilon_{13}, \varepsilon_{17}, \varepsilon_{21}$	remain On-Target in $x_2(t)$	$h_5(\mathbf{x}) < 0$	up
$\tilde{x}_{14}, \tilde{x}_{18}, \tilde{x}_{22}$	Off-Target in vertical $(x_2(t))$	$h_6(\mathbf{x}) < 0$	up
$\varepsilon_{14}, \varepsilon_{18}, \varepsilon_{22}$	remain Off-Target in $x_2(t)$	$h_6(\mathbf{x}) < 0$	down
$\tilde{x}_{15}, \tilde{x}_{19}, \tilde{x}_{23}$	On-Target in vertical $(x_2(t))$	$h_7(\mathbf{x}) > 0$	up
$\varepsilon_{15}, \varepsilon_{19}, \varepsilon_{23}$	remain On-Target in $x_2(t)$	$h_7(\mathbf{x}) < 0$	down
$\tilde{x}_{16}, \tilde{x}_{20}, \tilde{x}_{24}$	Off-Target in vertical $(x_2(t))$	$h_8(\mathbf{x}) < 0$	down
$\varepsilon_{16}, \varepsilon_{20}, \varepsilon_{24}$	remain Off-Target in $x_2(t)$	$h_8(\mathbf{x}) < 0$	up
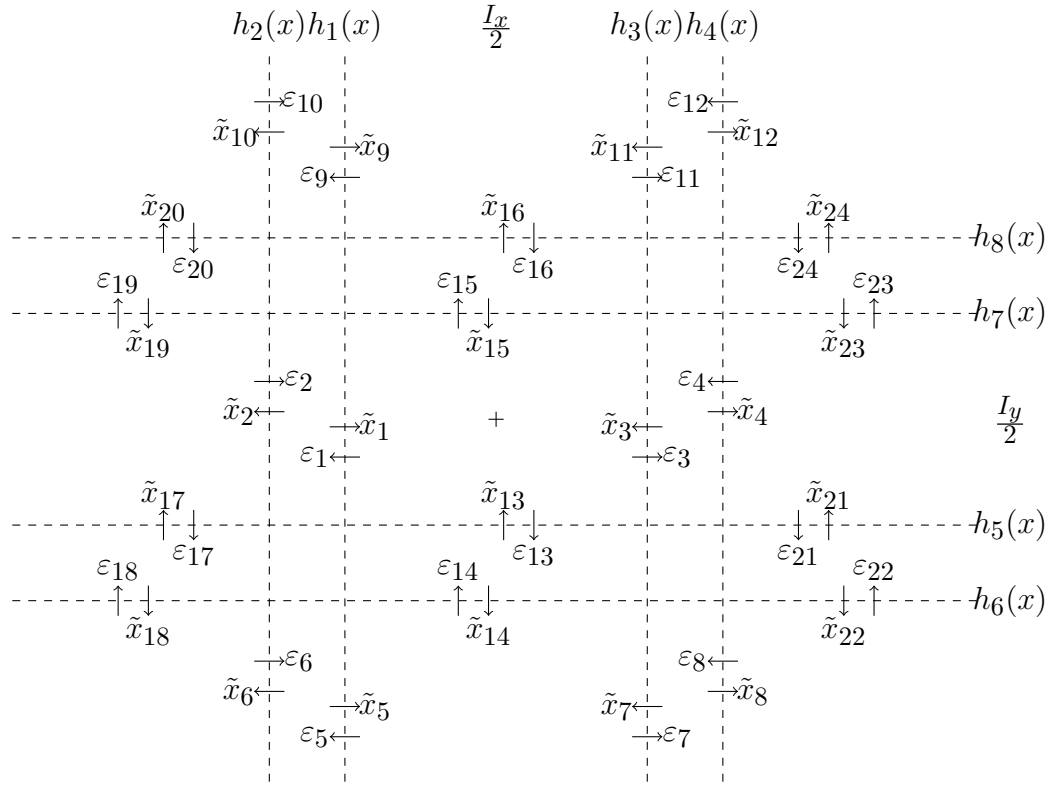
Figure 3.11: Illustration of the silent and non-silent events triggered by directional crossings of the hypersurface boundaries within the 2-D image image state space

ing the compensatory tracking solution in accord with the DES-controller output. The well-behaved transitions follow trajectories that are moving towards resolving the compensatory tracking solution not necessarily in accord with the DES-controller output. An example is the state transition $\tilde{s}_5 \rightarrow \tilde{s}_2$ on $\frac{\varepsilon_{18}}{\tilde{r}_2}$ which resolves the vertical component of the tracking solution even though the axially prioritized control law attempting to drive the plant towards resolving the horizontal component first. The ill-behaved transitions follow trajectories that are moving away from resolving the compensatory tracking solution in contradiction to the DES-controller outputs. An example is the state transition $\tilde{s}_7 \rightarrow \tilde{s}_7$ on $\frac{\varepsilon_9}{\tilde{r}_5}$ which moves outside the definite horizontal `On-Target` region subspace under a control law attempting to drive the plant to resolve the vertical component, erroneously resulting in a larger $||p_{xy}||$.
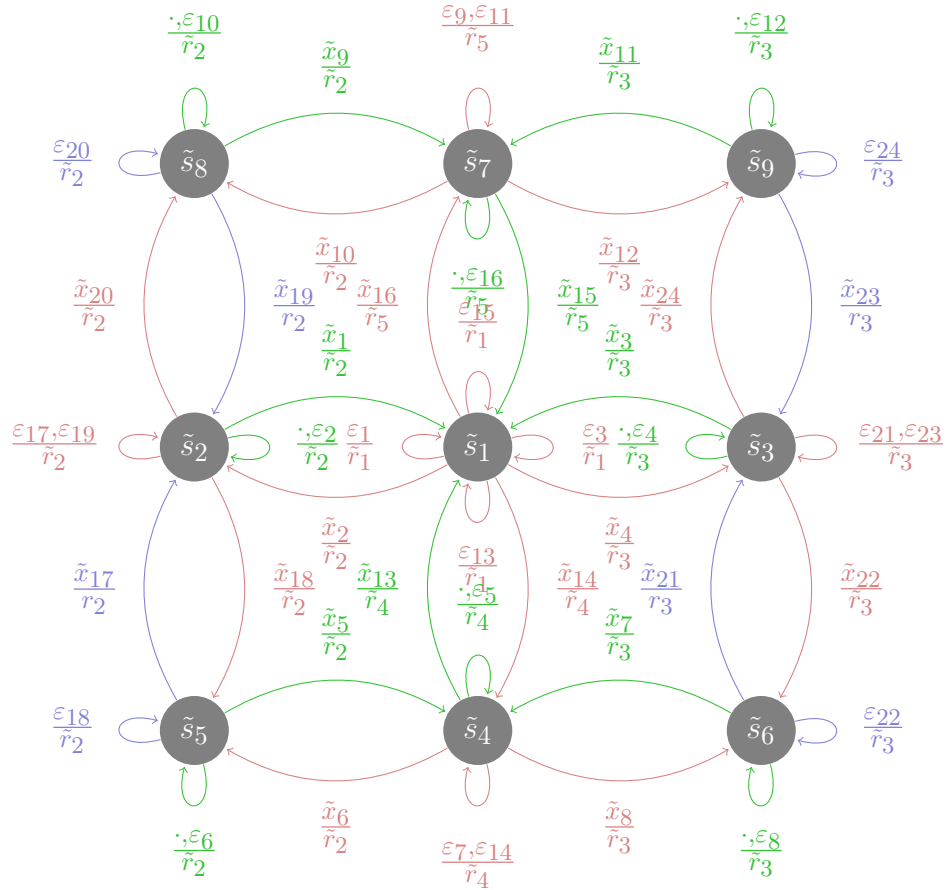
Figure 3.12: DES-controller state transition diagram for two dimensional l_2-model. Green indicates-ideally behaved transitions, blue indicates well-behaved transitions, and red indicates ill-behaved transitions

3.6.5   The 3-D DES-Controller Model with Hysteresis

Now we can expand to the three-dimensional task space, keeping in mind that the state-space is still defined within image feature-space. By extending the state-space to include a third position component of depth based on the image feature moment, $m_{00} = -x_3$, representative of the size of the target object projected within the image, we can create the boundary that triggers an `On-Target`condition after sufficient depth motion once horizontal and vertical alignment is achieved.

The set of hypersurfaces is then constructed by taking the eight partitioning functions specified in equations (3.42)-(3.49) and defining an additional one, given by

equation (3.50).

$$h_9(\mathbf{x}) = x_3 + f_z \tag{3.50}$$

To provide a simplified illustration of the DES-controller state transitions, we overlay the ideal set of state transitions upon an illustration of the hypersurface partitioning for the 2-D image-space in Figure 3.11. The depth boundary hypersurface, $h_9(\mathbf{x})$, is not shown but is located f_z into the page. The DES-controller state symbols are located within the regions of the state-space that are definite. Similar to the 2-D case above, the DES-controller will take one of two adjacent states within the regions between the hypersurfaces depending on the CT-plant's trajectory through the state-space, with the exception of depth. Once the end-effector initially crosses the depth boundary it is sufficiently close that the reaching task is complete. The user can take control of the overall task to probe the vicinity manually to locate the surface of the object and then further proceed to determine the appropriate grasping pose and forces.

The configuration of the green state transitions shows that horizontal compensatory tracking is prioritized over vertical. Horizontal deviations are resolved first within each image frame before any vertical motion cues are given. Similarly, vertical deviations are prioritized over depth deviations.

The DES-controller state, DES-controller output symbol, and DES-plant events are inherited from the two dimensional model with some minor additions and redefinitions are given in Table 3.5. Since the hypersurface boundary locations are given in pixels, the size of the physical target region bounded by their projection out into the task space will vary with the distance to the target. This means that the user is only required to exhibit precise control over the movement of their hand/camera when they are very close to the target object. At moderate distances, the optical axis could possibly be directed at some portion of the target object but not its centroid and still be `On-Target`. This method of partitioning prevents the user from unnecessarily expending energy attempting to attain and maintain precise registration at distances where small deviations in navigational trajectory are irrelevant to the task.

Depicting the entire state diagram is fairly cumbersome so the diagram in Figure 3.13 illustrates the transitions between in the subset of DES-Controller states for the 3D l_2 model localized in the lower left quadrant of the image frame. Through symmetry, the reader can envision the other state transitions from the remaining

Table 3.5: DES-plant events $(\tilde{X})$, DES-Controller state symbols $(\tilde{S})$, and DES-controller output symbols, $(\tilde{R})$ that extend the two dimensional l_2-model to three dimensional tracking

Symbols Redefined	Symbols Added
	$\tilde{x}_{25}$: On-Target in depth, $h_9(\mathbf{x}) > 0$
$\tilde{s}_1$ Off-Target-Depth	$\tilde{s}_{11}$ On-Target
$\tilde{s}_{10}$ No-Target	
$\phi(\tilde{s}_1) = \tilde{r}_1$: Move-Forward	$\phi(\tilde{s}_{11}) = \tilde{r}_7$: Stop

three quadrants of the image frame. The diagram is very similar to the lower left portion of Figure 3.12 with some distinct differences. The first is the depth tracking transition, $\tilde{s}_1 \rightarrow \tilde{s}_{11}$ on $\frac{\tilde{x}_{25}}{\tilde{r}_7}$, and the second is the loss of target transitions from the four intermediate tracking states to $\tilde{s}_{10}$.
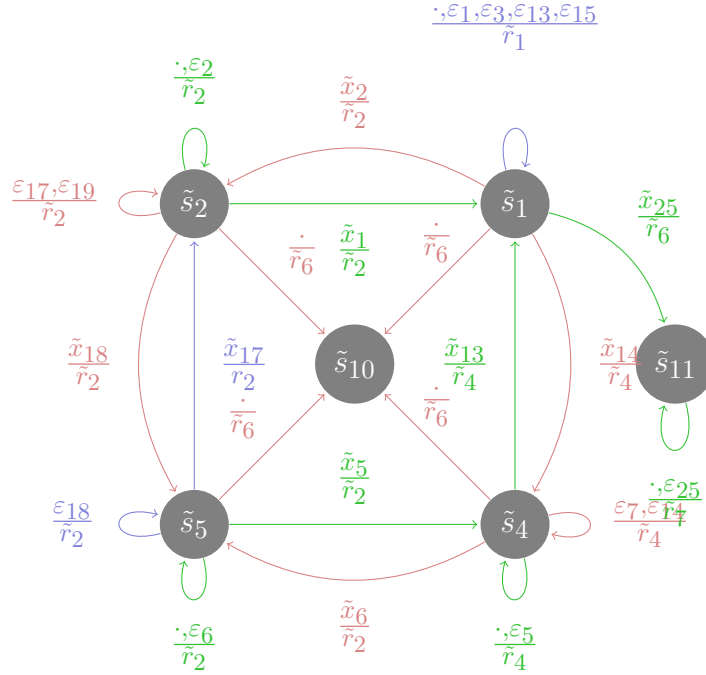
Figure 3.13: An illustration that depicts state transitions within the 3-D image state-space that traverse hypersurface boundaries $h_1(x)$ to $h_9(x)$

3.6.6 Continuous-Time Plant Input

The actuator function within the interface, $\gamma : \tilde{R} \to \mathbb{R}^m$, converts the DES-controller output symbols into a piecewise continuous input signal to the CT-plant. It is given by

$$\mathbf{r}(t) = \sum_{n=0}^{\infty} \gamma(\tilde{r}_n) I(t, \tau_c[n], \tau_c[n+1]) \tag{3.51}$$

where $I(t, \tau_c[n], \tau_c[n+1]) = 1$ over the interval $\tau_1 \leq t < \tau_2$, and $\tau_c[n]$ is the timestamp of the nth DES-controller output symbol, $\tilde{r}[n]$.

The use of vibratory signalling through the plant's somatosensation input channel allows for the implementation of either bang-bang and proportional control mechanisms. For bang-bang control a constant vibration is applied to the vibrotactor(s) regardless of the magnitude of error between the center of the frame and the centroid of the target. With a proportional scheme the frequency varies according to the

magnitude of the error. In either case, the the discrete automaton supervisor still switches amongst the set of one-dimensional control laws.

The horizontal and vertical controls $r_x(t)$ and $r_y(t)$ are of the same form for both the bang-bang and proportional schemes. The control for bang-bang $r_x(t)$ is given in equation (3.52) and the control for proportional $r_x(t)$ is given in equation (3.53).

$$r_x(t) = u(x_1 - l_1 I_x) - u(-x_1 - l_1 I_x) \tag{3.52}$$

$$r_x(t) = \frac{x_1 - l_1 I_x}{\frac{I_x}{2} - l_1 I_x} u(x_1 - l_1 I_x) - \frac{x_1 + l_1 I_x}{\frac{-I_x}{2} + l_1 I_x} u(-x_1 - l_1 I_x) \tag{3.53}$$

Given that l_1 and l_2 are in measured in pixels, the target region bounded by the hypersurfaces projected out into the task space will vary with the distance to the target. This means that the user is only required to exhibit precise position control over the movement of the camera when they are very close to the target object. At moderate distances, reasonable misalignment is allowable which simplifies the trajectory traversal.

The depth control $r_z(t)$ differs in that there is only a hypersurface at some point in front of the camera, i.e. the control only drives the plant in one direction. If the camera view is beyond the target, the system is unstable. Also, moderate tracking in depth can cause the DES-controller to switch to a horizontal (or vertical) tracking state if there was moderate target centroid misalignment at the onset of the depth movement requiring the new horizontal (or vertical) error to be resolved before the depth tracking state, $\tilde{s}_{11}$, can continue. Under a proportional scheme, upon a resuming $\tilde{s}_{11}$, $r(t)$ would be at the maximum in its frequency range creating an ambiguous message to user. In the real sequence of events, they had moved closer to the target, then resolved some minor horizontal (or vertical) misalignment, resulting in a depth tracking signal which indicates that their hand is now further away from the target. As such, only bang-bang control, $r_z(t)$, is used for depth track and is given in equation (3.54)

$$r_z(t) = u(x_3 + f_z) \tag{3.54}$$

3.7 Hybrid Control System Performance Measures

With the state-space partitioning models described in the previous sections, reaching task tracking performance can be quantified as discrete event measures that show the accuracy of response to the control issued and the precision of the response. Using either separately, or a combination of the two will provide varying degrees of success in the performance of the assisted reaching task and the behaviour of the plant.

In this section we revisit and expand upon the three behavioural classes of system performance that were introduced in Section 2.5. The classes are generally defined as:

Ideally-behaved indicates that the user followed the motion cues with a high level of precision and accuracy, continually decreasing the compensatory tracking distance;

Well-behaved indicates that the user followed the motion cues with a reasonable degree of precision and accuracy, generally decreasing the compensatory tracking distance;

Ill-behaved indicates that the user exhibited, cumulatively over the entire trajectory, a significant deviation from the motion cues issued.

In terms of an accuracy measure, traversal through the state-space during the reaching task triggers the events. The sequence of the events and the composition of the sequence can describe the plant's response to the control signal (motion cues) issued. The diagram in Figure 3.14 shows the regions of the partitioned image state-space in which each DES-controller state, $\tilde{s}_i$, $i = 1 \ldots 9$, operates. Starting within any particular region, the DES-controller will try to drive the plant through a trajectory that would produce a specific sequence of DES-plant events. Depending on initial start position in the state space an ideal trajectory would be described by a sequence of n symbols composed of one or unique DES-plant event symbols. The symbols can be divided into the axially prioritized subsets $\tilde{X}_1 = \{\tilde{x}_1, \tilde{x}_3, \tilde{x}_5, \tilde{x}_7, \tilde{x}_9, \tilde{x}_{11}\}$ for horizontal tracking, then the appropriate symbol from $\tilde{X}_2 = \{\tilde{x}_{13}, \tilde{x}_{15}\}$ for vertical tracking, and finally $\tilde{X}_3 = \{\tilde{x}_{25}\}$ for depth. Assuming two tasks have the same start and and target position in the task space, and each is performed in the the same level

of accuracy and precision, the length of the sequence, N, is dependent on end-effector velocity which is solely determined by the user. Thus basing a metric on directly on velocity profile or completion is not advantageous nor would it necessarily produce consistent results.
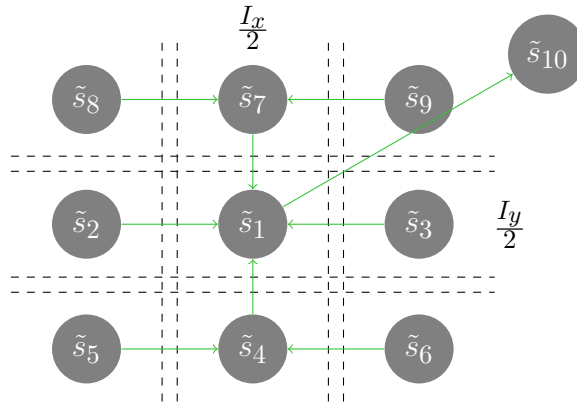
Figure 3.14: An illustration of the two dimensional state-space that depicts all the ideally-behaved DES-controller state transitions triggered from non-silent DES-plant events that traverse hypersurface boundries $h_1(x)$ to $h_8(x)$

Instead, we proposed a set of metrics consist of an accuracy measure, a precision measure, and an overall performance measure which is a function of both. These However, within the scope of that string of DES-plant event symbols, portions of the tracking trajectory could follow winding marginally stable paths within each of the partitioned regions through the state-space. As such, a minimal sequence of DES-plant events could be a sample path that is a less efficient completion of the reaching task than an alternate sequence with a more accurate track of the intended trajectory specified by the DES-controller output symbols. Thus an additional motion cue tracking performance measure which specifies precision is described in the next subsection. We use both accuracy and precision measures to examine the real trajectories recorded from subjects during reaching task experiments using the aiReach system.

3.7.1 Discrete Event Measures

While the sequence of DES-plant events triggered by traversal of the state-space drives and is driven by the evolution of DES-controller state transitions, it also describes

the accuracy of the DES-plant's response to the DES-controller output. Each DES-controller output symbol, resulting in a CT-plant input, if followed even with a broad degree of precision would produce a well-behaved sequence of subsequent plant events since DES-controller states operate over regions of the state-space.
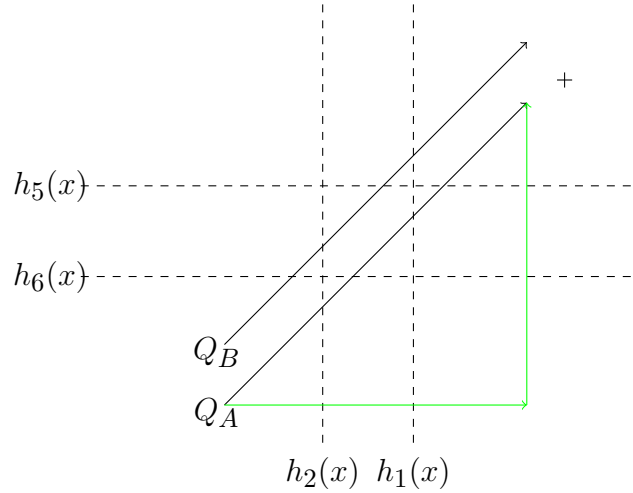
Figure 3.15: Diagram that shows two terminal point equivalent trajectories, Q_A and Q_B, originating from within the lower left quadrant of the image-space

The diagram in Figure 3.15 shows two possible alternate, parallel trajectories, Q_A and Q_B, occurring between times $\tau_e[n]$ and $\tau_e[n+k]$. Both originate in the region of the state-space that is driven under DES-Controller state $\tilde{s}_6$ and terminate in the region driven by $\tilde{s}_1$. Also included, in green, is the ideal trajectory that should evolve from the same initial point as Q_A if the plant (user) DES-plant response to the DES-Controller output symbols was highly accurate and precise. While each Q_A and Q_B cross the same four hypersurfaces, they do so in a differing order. Thus the events triggered and the corresponding state transitions will differ by producing the following plant events / controller states / output symbol sets:

Q_A: $\tilde{X}_A = \{\cdot, \varepsilon_6, \varepsilon_{18}, \tilde{x}_5, \tilde{x}_{13}\}$, $\tilde{S}_A = \{\tilde{s}_5, \tilde{s}_5, \tilde{s}_5, \tilde{s}_4, \tilde{s}_1\}$, and $\phi(\tilde{s})_A = \{\tilde{r}_2, \tilde{r}_2, \tilde{r}_2, \tilde{r}_4, \tilde{r}_1\}$

Q_B $\tilde{X}_B = \{\cdot, \varepsilon_{18}, \varepsilon_6, \tilde{x}_{17}, \tilde{x}_1\}$, $\tilde{S}_B = \{\tilde{s}_5, \tilde{s}_5, \tilde{s}_5, \tilde{s}_2, \tilde{s}_1\}$, and $\phi(\tilde{s})_B = \{\tilde{r}_2, \tilde{r}_2, \tilde{r}_2, \tilde{r}_2, \tilde{r}_1\}$

In terms of a region basis the accuracy of the two trajectories is equivalent. Yet examining the individual DES-controller state transitions that differ between the two trajectories shows that Q_A has only ideally-behaved transitions, whereas Q_B has one

well-behaved trajectory. That difference is at the third state transition in Q_B where $\tilde{s}_5 \rightarrow \tilde{s}_2$ on $\frac{\tilde{x}_{17}}{\tilde{r}_2}$ as the non-ideal transition. So we can define a measure to evaluate the accuracy behaviour of the trajectory evolution by evaluating each sequential state transition by

$$\delta[n] = m(\tilde{x}[n], \tilde{s}[n], \tilde{s}[n-1]) \tag{3.55}$$

and then generate an accuracy behaviour score for the entire trajectory throughout the reaching task by

$$D = \frac{1}{N-1} \sum_{n=1}^{N-1} m(\tilde{x}[n], \tilde{s}[n], \tilde{s}[n-1]) \tag{3.56}$$

with $m(\tilde{s}[n], \tilde{s}[n-1])$ being a discrete function the produces values from $\{1, 0, -1\}$ for state transitions that are ideally-behaved, well-behaved, or ill-behaved, respectively. Equation (3.55) produces the trajectory segment based sequence of state transition scores that is a term in the reaching task performance metric described later. The state transition metric, $-1 \leq D \leq 1$, given in equation (3.56) describes how accurately, on average, the plant reproduced the set of state transitions indicated by the DES-controller through the motion cues issued. A value of $D = 1$ indicates that the fully correct set of state transitions were followed during the trajectory. A value of $D = 0$ indicates that the sequence of state transitions replicates the equivalent of a random set of trajectory segment movements: some portion driven directly towards the target, some driven indirectly towards the target, and some portion directly away from the target. A value of $D = -1$ indicates that the trajectory followed was in opposition to the intend motion cues. While theoretically possible, scores of $D < 0$ are impractical as it would require an artificial termination of the reaching task at some arbitrary point, assuming an infinite field of view for the camera or at the point the target leaves the field of view, which would not be considered a successful execution of the reaching task. Returning to the example provided in Figure 3.15, the sample trajectories of Q_A and Q_B produce state transition accuracy scores of $D_A = 1$ and $D_B = 0.75$, respectively. With this measure, the state transitions that occur during the evolution of each trajectory can be quantified in terms of accuracy towards appropriate state transitions for the reaching task solution.

The DES-Controller state transitions can be filtered based on various subsets of silent events that occur during the evolution of a trajectory. The diagram in
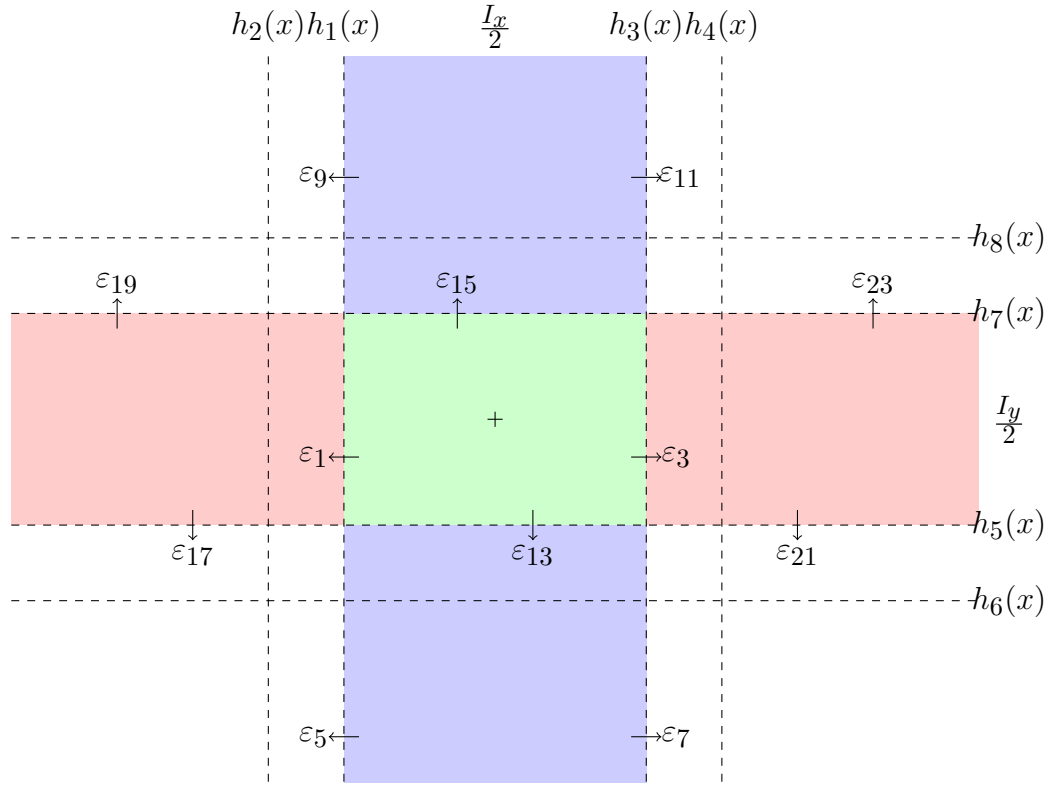
Figure 3.16: Region partitioning of the image space that shows silent events triggering equivalent state transitions.

Figure 3.16 shows three subsets of silent events: $\{\tilde{X}_\alpha\} = \{\tilde{\varepsilon}_1, \tilde{\varepsilon}_3, \tilde{\varepsilon}_{13}, \tilde{\varepsilon}_{15}\}$, $\{\tilde{X}_\beta\} = \{\tilde{\varepsilon}_5, \tilde{\varepsilon}_7, \tilde{\varepsilon}_9, \tilde{\varepsilon}_{11}\}$, and $\{\tilde{X}_\gamma\} = \{\tilde{\varepsilon}_{17}, \tilde{\varepsilon}_{19}, \tilde{\varepsilon}_{21}, \tilde{\varepsilon}_{23}\}$. To aid the reader in visualizing the boundaries, the corresponding regions of the state-space in which $\{\tilde{X}_\alpha\}$, $\{\tilde{X}_\beta\}$, and $\{\tilde{X}_\gamma\}$ occur are colour coded as green, blue, and red, respectively.

Figure 3.16 show the bounded regions that maintain. The subset $\{\tilde{\varepsilon}_1, \tilde{\varepsilon}_3, \tilde{\varepsilon}_{13}, \tilde{\varepsilon}_{15}\}$ under the l_2 hypersurface model causes the DES-Controller state to maintain a $\tilde{s}_1$ in comparison to l_1 hypersurface model while would trigger non-silent events and a corresponding DES-Controller transition to either one of $\{\tilde{s}_2, \tilde{s}_3, \tilde{s}_4, \tilde{s}_5\}$

3.7.2 Motion Cue Tracking Performance

The second element of measuring this system's performance is the precision with which the user tracks the intended (motion cues) as the trajectory evolves. Even though the time interval between DES-controller symbols ($[\tau_c[n], \tau_c[n+1])$ is ap-

proximately constant, with no control applied to the velocity and acceleration of end-effector movement, there is no reference displacement available each time the end-effector position changes to compare magnitudes. The DES-plant (user) autonomously determines the end-effector velocity throughout the course of the reaching task. There is also no available mechanism to maintain consistency of that control from one task to another, either upon the same user or across different users. However, what is consistent is the egocentric directional reference frame. Thus we choose to measure the directional precision of movement for each trajectory segment. To quantify the directional precision of tracking we defined $\mathbf{q}[n]\epsilon\mathbb{R}^3$ to be the the nth segment of the user's actual trajectory through the reaching task in image space. With the unit vectors in the direction of motion cues axes we can define three scalar quantities, $\rho_{x1}, \rho_{x2}, \rho_{x3}\epsilon\mathbb{R}$, for each trajectory segment given by

$$
\begin{aligned}
\rho_{x1}[n] &= \frac{\mathbf{q}[n] \cdot \hat{\mathbf{x}}_1}{||\mathbf{q}[n]||} \\
\rho_{x2}[n] &= \frac{\mathbf{q}[n] \cdot \hat{\mathbf{x}}_2}{||\mathbf{q}[n]||} \\
\rho_{x3}[n] &= \frac{\mathbf{q}[n] \cdot \hat{\mathbf{x}}_3}{||\mathbf{q}[n]||}
\end{aligned}
\tag{3.57}
$$

Each $\rho_x[n]$ gives the directional error fraction (DEF) within the range $[-1\ldots1]$ per trajectory segment. Similar to the equation (3.55), ρ_x is the segment based sequence of movement precision scores. The sequence is constructed from the appropriate $\rho_x[n]$ component corresponding to the current $\tilde{r}[n]$. Essentially it measures quality of the expended effort in the intended direction of motion. If $\rho_x[n] = 1$ then entirety of the displacement of the end-effector was along the intended direction of the motion cue, whereas $\rho_x[n] = 0$ would show orthogonal movement and $\rho_x[n] = -1$ would show opposing movement during the n-th trajectory segment.

To quantify the movement precision for the entire trajectory we define the directional tracking error as the average DEF, given by

$$
\theta_e = \frac{1}{N-1} \sum_{n=1}^{N-1} \rho_x[n]
\tag{3.58}
$$

Values of $-1 \leq \theta_e < 0$ indicate ill-behaved tracking since, on average, movements were

directed away from the intended direction indicated by the DES-controller output symbols. Values of $\theta_e = 0$ indicates that the movements were equivalent to an essentially random sequence, which is considered ill-behaved as there should be some bias towards movement corresponding to the motion cues given. Values of $0 < \theta_e \leq \beta$ are considered to be well-behaved and $\beta < \theta_e \leq 1$ are ideally-behaved. The threshold value of β allows for the realistic notion that ideally-behaved precision is not actually perfect. From a practical perspective, β is also likely to be unique to each individual and converges after successive sessions of practice.

The overall tracking response metric which quantifies performance of the reaching task is given by

$$\psi = \frac{1}{N-1} \sum_{n=1}^{N-1} \delta[n] \cdot \rho_x[n] \tag{3.59}$$

Referring back to the example depicted in Figure 3.15 we can then quantify the tracking response of the three sample trajectories as:

Q_A: $\delta[n] = \{1, 1, 1, 1\}$, $\rho_x[n] = \{0.707, 0.707, 0.707, 0.707\}$, and $\psi = 0.707$

Q_B $\delta[n] = \{1, 1, 0, 1\}$, $\rho_x[n] = \{0.707, 0.707, 0.707, 0.707\}$, and $\psi = 0.530$

Ideal: $\delta[n] = \{1, 1, 1, 1\}$, $\rho_x[n] = \{1, 1, 1, 1\}$, and $\psi = 1$

The scoring assignments above maintain the assumption that all motion is within the plane shown and the sample trajectories are subdivided into four segments, each only spanning one triggering event.

Chapter 4 Human Motor Control and Performance

Reaching task movements can be described in a number of different internal and/or external coordinate frames representing the trajectory of the hand from initial position and orientation to final position and orientation at the target. From a neurophysiological perspective, internal coordinate frames can be specified in terms of joint-space kinematics, joint-space dynamics, or vectorized patterns of actuating muscle activity at each joint [108]. Measurement of vector quantities for internal coordinate frame based control can range from moderately to extremely cumbersome (possibly invasive), and external coordinate frame representation is best suited for the work proposed here. We presented a set of proposed metrics in the later portion of the previous chapter derived from our hybrid control model for guidance of the reaching task. However there exists a well established metric within the context of natural human motor performance research which should be examined for completeness. Studies in human motor performance for pointing, or reaching tasks often employ Fitts' Law[32, 109] to quantify the degree of success relative to a subject's ability track in on a target. As a metric it quantifies an inherent speed-accuracy trade off that exists in human motor performance. The trade off manifests through a comparison between the predicted movement time required to complete the task and the difficulty of performing the task. The speed-accuracy trade off exhibits a linear proportionality between movement time and task difficulty. For the one dimensional case, the movement time, MT, is given by equation (4.1)

$$MT = a + b \cdot ID \qquad (4.1)$$

where ID is the index of difficulty for the reaching/pointing task, with a and b being experimentally derived constants. Fitts' Law uses an information theoretic approach to establish a linear relationship between the time required to preform the movement and the index of difficulty for that movement. The unit for ID, given in equation (4.2), is 'bits'. It shows that there is an inversely proportional relationship between

A, the amplitude of movement, and W, the target width along the line of approach when determining the index of difficulty for a particular movement.

$$ID = \log_2\left(\frac{2A}{W}\right) \tag{4.2}$$

The diagram in Figure 4.1 illustrates Fitts' classic reciprocal tapping test [32]. During the test, a subject is required to start at the midpoint of one target region, then move to and tap the corresponding target point within the opposing target region as quickly and accurately as possible. Upon the targeting tap, they are to continue with a reversal of direction and target the previous target region in successive back and forth motions. This difficulty of the a particular tapping test is determined by the ratio of movement amplitude between targets, A, and trying to tap within a region that is $\frac{W}{2}$ distance on either side of the target line.
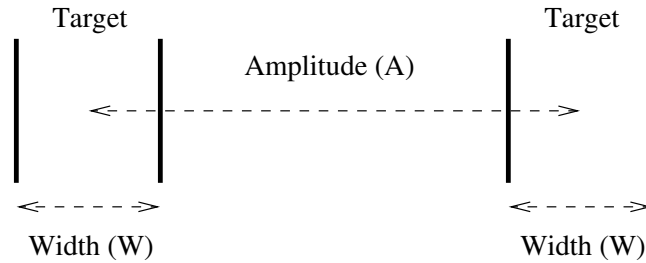
Figure 4.1: Illustration of the 1-D Fitts' reciprocal tapping test configuration

A number of variants to Fitts' Law have been proposed over the years. One that has gained wide spread adoption is the Shannon formulation proposed by Mackenzie [110] which is more closely related to Shannon's Theorem for the information capacity of a communications channel, given in equation (4.3),

$$C = B\log_2\left(\frac{S+N}{N}\right) \tag{4.3}$$

where the channel capacity, C, is a function of the signal power, S, and the noise power, N, given a channel bandwidth, B. The Shannon formulation of Fitts' Law is given by

$$MT = a + b\log_2\left(\frac{A}{W} + 1\right) \tag{4.4}$$

This formulation is preferred as it does not produce an erroneously negative value for ID when the amplitude of the movement is less than half the target width, nor an infinite value when the starting position for the task happens to be the target position as well ($A = 0$). The other notable formulation for targets constrained in one dimension was proposed by Welford [111] prior to MacKenzie's formulation, but is very similar.

The reader is referred to MacKenzie's paper [112] and a follow up work authored with Soukoreff [113], which provide an excellent review of the application of Fitts' Law as a quantitative performance model in the field of human-computer interfacing.

4.1 A Control Theoretic Approach

In this section we examine the approach of modeling the system performance from a control theoretic perspective in comparison to Fitts' information theoretic approach. A detailed discussion of the derivations presented in this section can be found in [114].

Continuing with the established premise that task difficulty corresponds to the relationship between the movement amplitude and the width of the target, we examine first- and second-order system models for movement performance.

Starting with a first-order system response to a step input, it can be shown to display a similar inversely proportional relationship between movement amplitude and target width. Using the same definitions for amplitude, A, and target width, W, given in equation (4.2). The ideal output position is the center of the target region; a movement amplitude of A from the start position. However, the target has actually been reached after traversing $A - \frac{W}{2}$ from the starting position. In the latter case the output has been driven to the acceptable value, the leading edge of target object. That is an acceptable target acquisition criteria and thus we get

$$A - \frac{W}{2} = A - Ae^{-kt}$$

Employing some algebraic manipulation and the change of base property for logarithms, we get the result

$$\frac{\ln 2}{k} \log_2 \left(\frac{2A}{W} \right) = t \tag{4.5}$$

which is essentially the original formulation of Fitts' Law with $a = 0$ and $b = \frac{\ln 2}{k}$.

Similarly, for a second-order system providing a mass-spring-dashpot model as the basis for the movement, the motion of the can be described by

$$F(t) - k_2 \dot{x}(t) - k_3 x(t) = \frac{1}{k_1} \ddot{x}(t) \tag{4.6}$$

where $F(t)$ is the input force, $\frac{1}{k_1}$ is the mass, k_2 and k_3 are the gains associated with the countering forces due to friction and the restorative force of the spring (co-activation), respectively. Since the system (human arm) is not inherently drawn back to its initial position we can use $k_3 = 1$ and obtain a transfer function of

$$\frac{x(s)}{F(s)} = \frac{\omega_n^2}{s_2 + 2\zeta\omega_n s + \omega_n^2} \tag{4.7}$$

which takes on the more common convention of the constant parameters being expressed in terms of the undamped natural frequency, ω_n, and the damping ratio, ζ.

In the context of a reaching/pointing task, Langolf *et al.* [115, 116] showed that response for the system is an underdamped solution, so the second order model produces the exponential form

$$A - \frac{W}{2} = A - A\left(\frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}}\right) \tag{4.8}$$

After employing similar algebraic manipulation, equation (4.8) can be expressed as

$$t = \frac{1}{-\zeta\omega_n} \ln(\sqrt{1 - \zeta^2}) + \frac{\ln 2}{\zeta\omega_n} \log_2 \frac{2A}{W} \tag{4.9}$$

In this form we get $a = \frac{1}{-\zeta\omega_n} \ln(\sqrt{1 - \zeta^2})$ and $b = \frac{\ln 2}{\zeta\omega_n}$. One of the key failures in using first- and second-order dynamic models to quantify reaching/point task performance is the mismatch between the theoretical and experimental velocity profiles of the movement. Data collected by numerous researchers [55, 50] shows that the velocity profiles for natural movement exhibits a gaussian (bell) shaped curve as opposed to the skewed, peakedness exhibited by the theoretical first- and second- order dynamic models. The application of the Fitts' law relationship as a performance measure describes the subjects capacity to perform an accurate targeting task as opposed

to attempting to describe that actual motion. In further contrast, the information theoretic approach of the Fitts law metric captures the variability of targeting in natural movement.

4.2   Fitts' Law Adaptations for 2-D and 3-D Targeting Tasks

Crossman [117] first proposed a two dimensional formulation of Fitts' Law shortly after Fitts published his initial work. The proposed model was based on only a pilot study using two subjects and took the form of

$$MT = a + b \cdot \log_2(\frac{A}{W} + 1) + c \cdot \log_2(\frac{A}{H} + 1) \qquad (4.10)$$

where a, b, and c are experimentally derived constants. The significance of Crossman's model is that it clearly delineates contributions to a movement by separating the difficulty from the amplitude resolution, $\frac{A}{W}$, and the difficulty from the directional resolution, $\frac{A}{H}$. MacKenzie and Buxton [118] are recognized for the two most widely accepted 2-D formulations of Fitts' Law: the *apparent-width* model and the *smaller-of* model.

The first is the more intuitive and an abstraction of a one-dimensional pointing task. The dimensional reduction is achieved by only considering the line of approach to the target, as shown in Figure 4.2, when determining the amplitude of movement and the target width, W', The formulation for the *apparent-width* model is given in equation (4.11). Even though there are only to independent quantities in the $ID_{W'}$ formulation, it is dependent on four parameters, A, W, H, and θ; which makes it slightly more complex than the *smaller-of* formulation given in equation (4.12), which is only dependent on the parameters A, W, and H.

$$ID_{W'} = \log_2\left(\frac{A}{W'} + 1\right) \qquad (4.11)$$

$$ID_{\min(W,H)} = \log_2\left(\frac{A}{\min(W,H)} + 1\right) \qquad (4.12)$$

They also compared three other formulations: $ID_{W \times H}$, ID_{W+H}, ID_W (contribution of H is ignored, referred to as *status-quo* model). Those three were discounted as viable models due to statistically significant differences in the pairwise comparison between the correlation of those models to the experimental data and correlations of the *apparent-width* and *smaller-of* models to the experimental data. Their results reported that while the *smaller-of* model showed a slightly stronger correlation to the experimental data than the *apparent-width* model, the pairwise comparison between the two was not statistically significant.

Hoffmann and Sheikh [119] also separately proposed an $ID_{\min}$ model for 2-D pointing task. In their work, they provided a justifying rationale over the Crossman model by arguing "only when the target height is less than the natural vertical scatter of hits on the target is there likely to be any effect of vertical constraint"(Hoffmann *et al.* , pg1073). This is a valuable insight as it prescribes a threshold for the onset of a directional pointing task component within the overall task. The onset threshold being the outer endpoints of the targeting scatter perpendicular to the line of approach.

For the purposes of simplicity in notation, from this point on we drop the prime superscript and adopt the convention that W extends in the direction of movement which relates to the amplitude task constraint, and H extends orthogonal to the direction of movement which relates to the directional task constraint.

Other researchers such as Ware and Balakrishnan [120] and Murata [34, 35] used the Shannon formulation in subsequent work. However, a study of bivariate pointing tasks by Accot and Zhai [121, 122] showed that there were fundamental limitations with the both the $ID_{\min}$ and $ID_{W'}$ models. One of those limitations is the inconsistent interaction between the contributions to MT as W and H vary. At the extreme, when either $W \to \infty$ or $H \to \infty$, both the $ID_{\min}$ and Crossman models adequately represent either an exclusively directional or exclusively amplitude pointing task, respectively. However, within the range of nominal values for W and H the $ID_{\min}$ model does not adequately capture the interaction between the two dimensions. The model predicts that the value of H is irrelevant to the task difficulty if $H > W$ and $\frac{H}{W} \approx 1$. Similarly, it predicts that W is irrelevant to the task difficulty if $W > H$ and $\frac{W}{H} \approx 1$.

Accot and Zhai also discussed a number of properties missing from the current models in the literature that are necessary to more accurately capture the behaviour

of a 2-D Fitts pointing task:

- movement time scale independence when A, W, H are all multiplied by the same factor;

- regression to a 1-D Fitts model as either W or H goes to infinity;

- the smaller of either W or H should be the dominant factor in the index of difficulty;

- the duality in the nature of the effect of W and H on the index of difficulty;

- the effect of W and H should be continuous over the range of values.

Accot and Zhai proposed a new 2-D formulation based on vector norms which is described below. Given a vector $\mathbf{x} = (x_1, \ldots, x_n)$, a set of weights $\mathbf{w} = (w_1, \ldots, w_n)$, and $p \epsilon \mathbb{R}$ the weighted ℓ_p-norm of x is defined by:

$$\| x \|_{p,w} = \left(\sum_{i=1}^{n} w_i \mid x_i \mid^p \right)^{\frac{1}{p}} \tag{4.13}$$

Applying it to the pointing task constraint vector $X = (\frac{A}{W}, \frac{A}{H})$ results in a bivariate model for a 2-D Fitts pointing task of the form

$$T = a + b \log_2(\| X \|_{p,w} + 1) \tag{4.14}$$

and allows the scaled contribution of both target dimensions in the calculation of the index of difficulty. In that fairly comprehensive study, they proposed and compared three formulations of a bivariate 2-D Fitts Law based on the ℓ_1-norm , ℓ_2-norm, and ℓ_∞-norm of the constraint vector, $X_{p,w}$, in the index of difficulty:

$$ID_{\ell_1} = \log_2 \left(w_1 \frac{A}{W} + w_2 \frac{A}{H} + 1 \right) \tag{4.15}$$

$$ID_{\ell_2} = \log_2 \left(\sqrt{w_1 \left(\frac{A}{W} \right)^2 + w_2 \left(\frac{A}{H} \right)^2} + 1 \right) \tag{4.16}$$

$$ID_{\ell_\infty} = \log_2 \left(\max \left(w_1 \frac{A}{W}, w_2 \frac{A}{W} \right) + 1 \right) \tag{4.17}$$

The ℓ_∞-norm with unary weights can be rewritten as $\frac{A}{\min(W,H)}$ which represented the prior state of the art in 2-D Fitts Law formulations according to MacKenzie and Buxton [118], as well as Hoffmann and Sheikh [119]. Their results showed a significant difference in the effect on movement time dependent on the ratio of the target dimensions. When the targeting task is amplitude dominant ($\frac{H}{W}$, $W \leq H$) then MT is essentially constant as H decreases from infinity to W, as opposed to a directionally dominant targeting task ($\frac{W}{H}$, $H \leq W$) which shows an approximately linear decrease in MT as W decreases from infinity to H. Using both their experimental data and the raw data from [119], Accot and Zhai were showed that the weighted Euclidean formulation given in equation (4.18) was the best fit to the available experimental data.

$$T = a + b \log_2 \left(\sqrt{\left(\frac{A}{W}\right)^2 + \eta \left(\frac{A}{H}\right)^2 + 1} \right) \tag{4.18}$$

Of key note is the experimentally determined weighting factor, η, on the directional constraint term; allowing the model to more accurately capture the interdependence of target dimensions and task amplitude on predicted movement time.

Other extended derivations of Fitt's Law that apply to 2-D and 3-D targeting tasks have been proposed by [33, 34, 35, 36, 37] and others. In particular, Grossman *et al.* [37] followed up on Accot and Zhai's work by investigating 3-D variants of the ID_{ℓ_2} and ID_{ℓ_∞} formulations. Those formulations are given in equations (4.19) and (4.20), respectively.

$$ID_{\ell_2} = \log_2 \left(\sqrt{f_W(\theta)(\frac{A}{W})^2 + f_H(\theta)(\frac{A}{H})^2 + f_D(\theta)(\frac{A}{D})^2 + 1} \right) \tag{4.19}$$

$$ID_{\ell_\infty} = \log_2 \left(\frac{A}{\min(f_W(\theta) \cdot W, f_H(\theta) \cdot H, f_D(\theta) \cdot D)} + 1 \right) \tag{4.20}$$

Where weighting parameters are a function of the movement angle, $f_{W,H,D}(\theta)$, towards the target. Within their study, they limited pointing task trajectories to a plane parallel to the transverse plane and used a fixed approach angle of zero so that approach was parallel to the target width dimension.

4.2.1 Fitts Law Extended to Trajectory Tracking or Steering Tasks

One of the fundamental aspects of the Fitts Law relationship is that the target size constraint is at the terminal point of the movement task. In general, no bound is placed on the intermediary trajectory along the amplitude until crossing the initial edge of the target. Accot and Zhai [123] in an earlier study chose to investigate the application of Fitts Law to trajectory-based tasks that required steering the end-effector through a "tunnel" region where the was a directional constraint on the path along the entire amplitude of the movement. To develop and validate a trajectory-based task the authors first used an experimental setup similar to the standard Fitts tapping test but with some alterations to the protocol. Subjects started from an initial position outside the bounds of the tunnel. Employing one-way discrete movements, the end-effector was tracked as it passed within a given height[1], H, perpendicular to the intended path at the start of the tunnel, then along the straight-line trajectory of amplitude A until it crossed the end of the tunnel within an identical terminal height constraint. Given a height constraint at both the initial point and the terminal point, Accot *et al.* reclassified this as a two-goal passing task. They verified that the two-goal passing task had a log-linear relational between A, H, and MT, just as Fitts Law and then further extended the model to a $N + 1$-goal passing task with each success goal being H in height and $\frac{A}{N}$ further along the path. This generates a index of difficulty, $ID_N = N \log_2(\frac{A}{NH} + 1)$. As $\lim_{N \to \infty} ID_N = \frac{A}{H \ln 2}$. Thus producing a linear-linear relationship between A, H, MT so that

$$MT = a + b\frac{A}{H} \tag{4.21}$$

Using thirteen subjects in a fully-crossed, within-subjects factorial design incorporating four amplitudes and eight tunnel heights the authors were able to get strong agreement between their model and the experimental data captured. The results produced a linear fit of $MT = -188 + 78 \cdot ID$ with $r^2 = 0.968$, and average

---

1. The authors of [123] used the terminology of width (W) for the constrained size of the tunnel boundary, but that is in conflict with the convention terminology used in Fitts Law studies. Width (W) refers to target dimension along the line of approach not perpendicular to it. We have adjusted the terminology here, and use "height" (H) where appropriate to avoid confusion.

error rate of 6.4%.

Friedlander *et al.* [124] similarly proposed the use of a linear-linear model relationship for their non-visual user interface element, called a *bullseye menu.* The example configuration is shown in Figure 4.3, and is essentially a goal-passing task similar to what was described above. While the target(s) are two dimensional, the height constraint is not strictly imposed given that it expands radially within each sector. The segment widths bounding each submenu item region within the sector are constant. Friendlander *et al.* conducted a series of experiments to determine employing a Fitts performance model or an alternative linear model for this type of user interface element. Deriving $A = r(x - 0.5)$ as the amplitude of movement for target selection of a particular menu item ring, with r being the width of each ring and x being the index for each ring. The Fitts Law model for a bullseye menu is then

$$MT = a + b \cdot \log_2(x + 0.5) \tag{4.22}$$

in comparison to the linear model which is

$$MT = a + b \cdot x \tag{4.23}$$

The authors collected movement time experimental data from 12 subjects performing a menu item selection task through 2208 trials over four sessions, and found better agreement between the data and the linear model that with the Fitts model. One of their key rationales for this result stems from the fundamental difference in targeting feedback loop. With a non-visual stimulus[2], there is a greater sensitivity to the large amplitudes of required movement.

4.2.2 Application of Fitts' Law to Non-sighted Reaching Tasks

Most formulations of Fitts' law are given in polar form assuming that the subject performing the reaching task will move their hand along a direct vector from the initial point to the target. This is a natural consequence of investigating pointing/reaching tasks undertaken by sighted individuals. We propose to examine a that validity

---

2. Friendlander *et al.* tested both tactile and auditory cues for signalling the index of each menu ring crossed

of a Fitts' law based performance measure for this visually assisted reaching task undertaken by non-sighted subjects.

Given that it is not feasible to issue motion cues that require the user to track precise joint angles; the motion cues must follow some easily referenced directions such as the axial components of an egocentric reference frame. This Fitts' law based measure could be created from the sum of one dimensional Fitts' law for each horizontal and vertical axial movement subtasks and Steering law for the depth axial movement subtask. Thus the total predicted time MT_T to complete the reaching task is given by

$$MT_T = MT_{x_1} + MT_{x_2} + MT_{x_3} \tag{4.24}$$

where are MT_{x_1} and MT_{x_2} are the expected subtask completion times to resolve the motions cues given by equation (2.3) and MT_{x_3} is the expected subtask completion time to resolve the motion cues related to the appropriate depth estimation technique for the various feature extraction techniques presented in Chapter 3.

4.3   Additional Relevant Literature

One of the notable issues raised in the literature about the appropriate application of Fitts' Law centers on the type of movement style in use: discrete or cyclical movements. There tends to be significant differences in the dynamics and perceptual-motor planning when the type of movement task only requires exerted movement in one direction towards a target as opposed to at least one return in the opposite direction to a reciprocal target set. Fitts' original experiments were conducting by subjects performing a reciprocal tapping test which employed a cyclical movement back and forth between the two terminal points (target strips).

Buchanan *et al.* [125, 126] performed a set of studies that investigated the change in dynamics of the end-effector as the harmonicity of the reciprocal tapping changed, driven by a change in the ID for the task. They systematically altered the target width from small ($ID = 5.85$) to large ($ID = 2.85$) and vice versa to determine the point of transition from harmonic (cyclical) to inharmonic (successive discrete) movement paradigms. Working from Guiard's [127] prior work which determined that reciprocal aiming task movements were harmonic when $ID > 4$ and inharmonic when $ID < 4$, Buchanan *et al.* chose to vary ID during trials every four seconds.

In one subset of trials ID was altered from large to small and in the other subset ID was altered from small to large, with amplitude remaining constant. They found that as target width varied from small to large the movement style transitioned from discrete to cyclical at $ID = 3.04$, well below the critical boundary ($ID = 4$). In contrast, when the varying target with from large to small, the cyclical to discrete movement transition occurred close more abruptly and close to the critical boundary at $ID = 3.53$. Their results show a hysteresis in the transition between harmonic and inharmonic movements depending on the initial style of movement. In short, motor planning for discrete movements has greater persistence in the presence of repetitious, reciprocal motion than cyclical harmonic movements. So in the case of the end-effector oscillation back and forth across a hypersurface boundary will likely continue to be a series of discrete movements as opposed to evolving into a transient cyclical movement.

4.3.1 Postural Issue

Almost the entirety of Fitts' law related studies of pointing/reaching/aiming based arm movement tasks are conducted with a range of postural constraints. Subjects are usually seated and in some cases arm movements are physically bound to a particular plane through an experimental rig used for measurement and data acquisition. The experimental rig can range from an affixed splint to a standard computer mouse or stylus pen. Bonnetblanc *et al.* [128] study investigating the effect on reaching/pointing task performance of subjects operating from a full upright standing position, similar to the postural state of user's of the aiReach system. They found that hand movement still exhibited Fitts' law performance in relation to changes in target size. Their results also showed a correlated increase in the durations of both the acceleration and deceleration phases and decrease in peak velocity as target size decreased, which indicates slightly more restrained movement dynamics in the formulated motor plan.
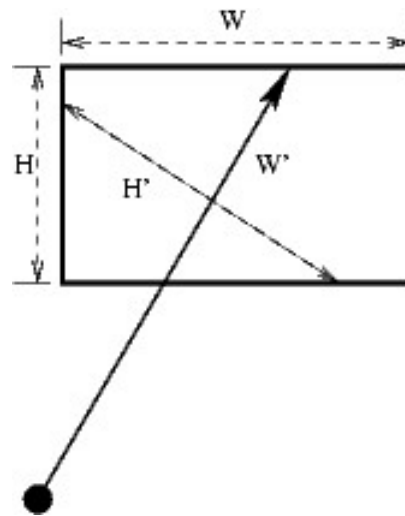
Figure 4.2: The alternate target with W' measured across the target object along the line of approach
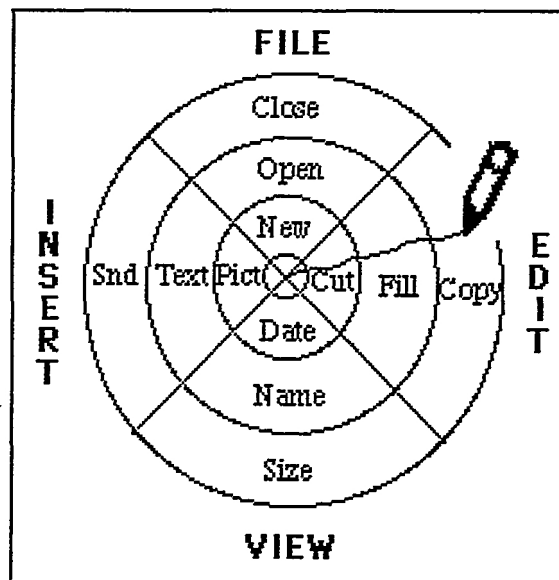
Figure 1: Selection from a bullseye menu.

Figure 4.3: An example of a bullseye menu proposed by Friedlander *et al.* . The image is a reproduction from *"Selection from a bullseye menu"*

Chapter 5 Description of Experiments

This chapter contains a detailed description of experimentation and testing undertaken during the development and verification of the *ai*Reach system. Analysis of usability testing, parameter tuning, and performance measurements are provided across the various sections describing the separate investigations performed.

5.1 Experiment 1: Prototype Construction – Proof of Concept

As discussed in section 1.1, a number of researchers have published results related to various designs for locomotive, navigational assistive devices for the visually impaired, but the specific task of a guidance aid for a reaching task is largely under-investigated. The initial task of constructing a basic working prototype was necessary to identify relevant testing factors and conditions.

The control architecture of the initial prototype implemented the 3D DES-Controller model without hysteresis (l_1 parameterization). That specific parameterization was not specifically detailed in Chapter 3, but is a simple extension of the material presented in section 3.6.2 with the addition of a fifth hypersurface delineating the state transition from a depth tracking state to the `On-Target` state within the DES-Controller.

This initial prototype was implemented using Bang-Bang control through the vibrotactile input to the CT-plant. The resultant CT-control signal generated while in the appropriate DES-controller state using equations (3.2), (3.51), and the appropriate hypersurface boundary functionals are:

$$
\begin{aligned}
r_{x_1}(t) &= u(\frac{I_x}{2}(1 + l_1) - m_{10}) - u(m_{10} - \frac{I_x}{2}(1 - l_1)) \\
r_{x_2}(t) &= u(\frac{I_y}{2}(1 + l_1) - m_{01}) - u(m_{01} - \frac{I_y}{2}(1 - l_1)) \\
r_{x_3}(t) &= -u(m_{00} + \alpha I_x I_y)
\end{aligned}
$$

where $u(\cdot)$ is the heavyside step function.

As an initial prototype, the size of the approximate target region was chosen by arbitrarily setting $l_1 = 0.1$. The value of $\alpha = 0.65$ was calibrated empirically and $I_x = 320$, $I_y = 240$ are taken from the resolution specifications of the particular camera used. A picture of the initial prototype is given in figure 5.1.
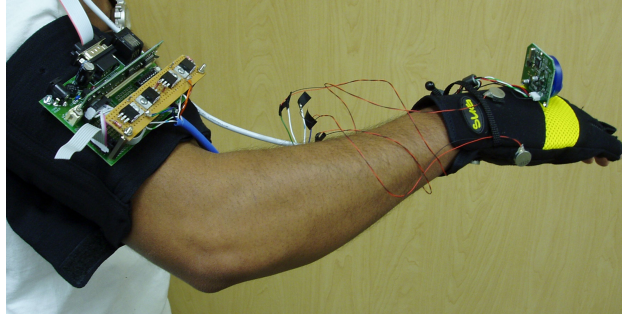
Figure 5.1: The initial prototype of the *aiReach* glove system

5.1.1 Experimental Apparatus

Initial development and testing of the feature extraction and object detection algorithm for simple planar geometric objects, described previously in section 3.2, was done using a pan-tilt servoing unit from Directed Perception Inc.[1] as the plant apparatus to verify the visual servoing of the basic 2D compensatory tracking task.

The wearable portion of the prototype: glove-mounted camera, tactile output transducers (vibrotactors), and controller board were assembled using readily available, inexpensive components. Both USB1.0 and IEEE1394 based cameras were tested.

The glove-mounted camera is connected to a PC running Windows with custom object tracking software written using the OpenCV library. The glove is also equipped with four vibrotactors; one each on the ventral (palm) and dorsal (back) parts of the hand, and either side. The placement of the four motors corresponds to the intended direction of motion of the user's hand within the x-y image plane; corresponding to horizontal and vertical movements within the task space. The actuation of the

---

1. http://www.dperception.com/

vibrotactors issuing the motion cues was controlled by the μCsimm[2] single board computer from Arcturus Networks; connected to the PC via a RS232 serial cable. The *mu*Csimm is a Motorola MC68EZ328 (Dragonball EZ) based single board computer with a port of the Linux 2.0.38 kernel known as μClinux[3]. Custom control software was written for the μCsimm to drive the vibrotactors based on tracking data sent from the object tracking application, specifying the desired hand trajectory.

The glove selected was a sports-utility glove with open finger tips. It was deemed appropriate that the user's finger tips should be exposed, so as to not restrict a visually impaired person's sense of touch. A sports-utility glove designed for use during physical activity also provided a durable and sufficiently padded construction to dampen some of the forces imparted by the vibrotactors. This allowed for comfortable use of the vibrotactile interface.

5.1.2 Experimental Procedure

Testing of the initial prototype is depicted in Figure 5.2. The environmental conditions for testing were not tightly constrained so that a reasonable approximation for real world conditions were used. Other than using planar geometric shapes of nearly uniform colour saturation on a neutral background, the only other environmental constraint was an attempt to maintain uniform ambient lighting on the target object field during tracking.

As can be seen in Figure 5.2, an assortment of planar geometric shapes of different colours are arranged at random on a neutral background; one of which is the target object. For these trials a red square was preset as the target object. Upon processing of each image frame, a list of candidate target objects within the scene is generated. Only closed polygons within the image frame are considered candidates. For each of the k candidate objects found, a coefficient of matching, M, is calculated based on the weighted sum described previously in equation (3.1) and show again below for convenience.

$$M = \min_k \sum_{j=0}^{n} (w_j(fd_{kj} - ft_j)^2)$$

---

2. http://www.arcturusnetworks.com/ucsimm.dragonball.ez.shtml

3. http://www.uclinux.org
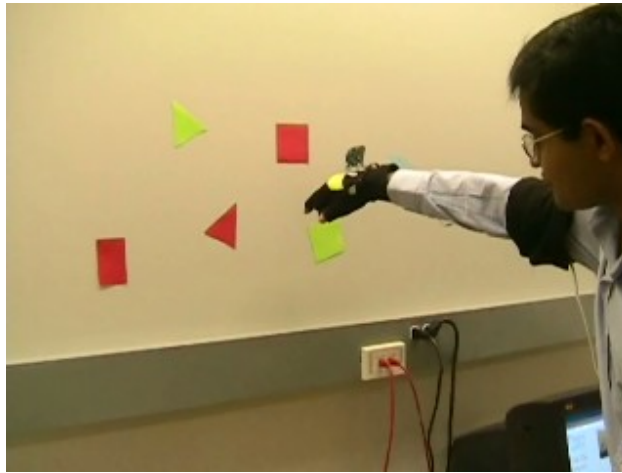
Figure 5.2: The initial prototype of the *aiReach* glove system during testing. This image is a still frame exported from a video recording of one of the initial experimental trials during the proof of concept testing of the prototype device.

The feature vector of length n associated with each of the k candidate targets contains elements describing basic geometric and colour space properties defined by the object model. For the purposes of this experiment those parameters include number of vertices, angle at each vertex, image coordinate of each vertex, colour value bounded within the vertices, etc. The associated weights, w_j, for each feature were set to strongly favour colour and allow for minor variations in vertex angle about a desired value of 90 degrees. The range of acceptable vertex angles mitigated prospective projection issues caused by the camera axis not being precisely perpendicular to the background.

For each trial, subjects started from an random initial distance from the target field with their eyes closed and arm extended in front of them in a comfortable posture, with the camera was directed toward the field of candidate target objects. Subjects were instructed to follow the directional motion cues until the vibrotactile symbol STOP was issued. After which they were to assume independent control for the remainder of the task and continue movement until contact with the target was achieved.

After initial system development and testing within our laboratory with five users, the initial prototype was demonstrated at a showcase booth at the 14th Annual Canadian Conference on Intelligent Systems. This was the prototype's first exposure

to untrained users outside the laboratory, as passers-by were encouraged to test the system in the unstructured setting of the demonstration booth at Conference's technology showcase. More than twenty volunteers participated in unstructured testing of the system and provided generalized qualitative feedback on the usability of the prototype.

5.1.3 Results and Discussion

After only a brief training period, nearly all users were able to successfully reach the target object during a trial by following the motion cues issued through the vibrotactors interface. Those who did exhibited a failed attempt did so by not completing the guided reaching task voluntarily. The unstructured testing, while encouraging and fundamentally proving the soundness of the design concept, demonstrated a number of issues within the system.

The variable lighting level across the background can be caused by a number of factors in an uncontrolled environment. For instance, shadows being cast by the user on the trial scene, or the unanticipated colour temperature of the lighting within the environment. Such variable lighting conditions, causing false positives in object detection, were a significant implementation issue. In addition, even though co-located objects were eliminated from the candidate list, constellations of several objects could be incorrectly detected as an single object.

With only an estimate of the typical size of the object of interest being used to eliminate this effect, the lack of a measured value for distance from the camera to the object plane, (denoted as the z-axis) was a problem with making an accurate decision based on size.

In the initial testing, with the camera is mounted on a precision pan-tilt device, two images acquired at two different poses could be used to estimate z, provided three or more landmarks registered as matched points within the two images. However, for the wearable system, this 'initialization' phase was not necessarily appropriate.

One of the initial prototype configurations used an IEEE1394 based camera that provided an image resolution of 1024x768. The overall system performance was not satisfactory in terms of tracking frame rate (refer to table 5.1). It was determined that to reduce the processing time, either we needed to adopt the use of a lower resolution image or reduce the complexity of processing by imposing constraints on

scene complexity. After further investigation, both options were implemented by first reducing the scene complexity with the use of a background that was nearly uniform in tone, then by switching to a USB1.0 webcam which had an maximum resolution of 320x240.

It was the lower resolution USB1.0 version that was demonstrated at 14th Annual Intelligent Systems Conference in Ottawa, Canada.

Table 5.1: Mean processing times in milliseconds (ms) for various stages of the image analysis software for both the IEEE1394 camera and USB1.0 camera

Camera	Image Capture	Object Detection	Control Signalling	Total Duration
IEEE1394	0.0 ms	738.0 ms	200.1 ms	938.1 ms
USB1.0	224.3 ms	32.0 ms	188.7 ms	445.0 ms

Two ergonomic issues became apparent during the testing. The first is related to the design choices involve in the selection of a glove. The thickness of the padding around the glove became a tradeoff between comfort and perceived tactile signal magnitude strength since some users, especially those with smaller hands, commented on their difficulty with feeling a distinct vibration pattern, localized to a particular vibrotactor.

When the object of interest is not present in the image, the present implementation of the system fails. The solution for this problem is to include an object search phase. A systematic search phase is easily implemented in a system under programmed control, however it is a much more complex task to accomplish for the wearable system since the human user is responsible for the high-level planning, and would have to be responsible for any "pan" to search the area.

A related observation during testing is the diversity joint position and joint trajectory configurations that evolve through a given reaching task trail. The joint configuration can vary significantly from initial to terminal position within the task space. For instance, if the user starts with an initial *Straight-arm* configuration, the evolution of joint angles at the wrist and the elbow exhibits almost no change. The vast majority of changes in joint angles occurs in the DoFs allocated to the shoulder joint(s). In contrast, if the user's initial position is some variation of a *Bent-arm* configuration, the evolution of joint angles and joint velocities can change consider-

ably across all the joint DoFs. This observed behaviour led to the development of Experiment 5.2.

5.2 Experiment 2: Static Loading Characterization

The DES-Controller model described in Section 3.6 specifies a parameter, l_1, used to position the hypersurface boundaries $h_1(x)$, $h_3(x)$, $h_5(x)$, and $h_7(x)$. The region enclosed by those boundaries delineates entry into the approximate target region. Specifying an approximate target region compensates for the user's inability to maintain a precise `On-Target` registration in image-space. It is necessary to investigate the natural motor stability of the user's arm under static load to determine an appropriate range of l_1 for steady-state `On-Target` registration and further compensatory tracking experimentation.

The purpose of this experiment is to determine the minimum value of l_1 that satisfies the behaviour of a generalized user when their arm is under a static loading condition. In the context of this work, a static loading condition is defined as the subject's attempt to maintain a constant targeting registration of the reaching arm in relation to an initial (`On-Target`) image-space location, for an extended duration. From qualitative observation of user behaviour during initial prototype testing, a range of nominal arm pose configurations was determined. However, it is unknown whether differences in arm pose significantly affect target registration stability.

Testing tracking performance at the relatively large or small values of l_1 can bias performance data by forcing a state transition prior to a targeting accuracy bound that the user is capable of, and cause a higher number of parallax induced state transitions during depth tracking. Thus this experiment was also designed to calibrate the useful range of values for l_1 in further experimentation relate to tracking performance.

5.2.1 Experimental Apparatus

The experimental apparatus consists of only the glove-mounted camera segment of the wearable portion of the aiReach system connected to the tracking control software. A simple spherical object is used as the target and is mounted on a shelving rig. The tracking control software was implemented based on the feature set model described in Section 3.4 to produce sufficiently high frame rate during video processing. The

image processing and feature extraction was implemented using a the RoboRealm[4] computer vision software package to create a RoboRealm application which executes a series of filter objects as an image processing pipeline to perform image acquisition, noise removal, colour segmentation, blob detection, and feature extraction for each frame captured. The pipeline is executed upon request from the control application via a TCP/IP socket connection. The extracted features correspond to feature vector, $\mathbf{f}_i = \begin{bmatrix} p_x & p_y & f_z \end{bmatrix}^T$, as described in Section 2.4. The control application records all image plane movement of the subject's hand by logging the position of the centroid, $\mathbf{p}_{xy} = [m_{10} \ m_{01}]$, of the target within the image.

The specific list of filters, in pipelined order, and their parameter values are given in Table 5.2. It is included for the purpose of replication of experiments. The parameter values within the camera properties object and RGB filter object are tuned, through trial and error, for the specific physical camera and ambient lighting conditions and environment within our laboratory where the experiment was conducted. Use with an alternate camera or under differing lighting conditions would benefit from manual calibration and parameter tuning before use of the aiReach prototype. The Blob_Size parameter is dependent on both the camera and the physical target object dimensions selected.

---

4. Available at www.roborealm.com.

Table 5.2: Configuration of filter parameters within the RoboRealm image processing pipeline.

Filter	Parameter	Value
Camera_Properties	brightness_value	5447
	contrast_value	3984
	software_auto_exposure	FALSE
	saturation_value	6748
	video_rate	30
	sharp_value	5041
	gamma_value	1512
	video_size	320 x 240
	whitebalance_value	5528
	video_format	RGB 24
	hue_value	4715
RGB_Filter	channel	2
	max_value	120
	hysteresis	5
	result_type	1
	min_value	85
Median	filter_size	7
Blob_Size	cutoff	30
	limit	1
	min_area	100
	object_size	10
	mask	FALSE
	max_area	60000
Smooth_Hull	window_size	7
Blob_Replace	shape_index	6
	color_index	4
	fill_shape	TRUE
Center_of_Gravity	show_coord	TRUE
	color_index	7
	connect_line	TRUE
	size_index	4
	density	-1
	use_subpixel	FALSE
	show_box	TRUE
	shape_index	2
	show_cog	TRUE
	threshold	-1

5.2.2 Experimental Procedure

The experiment employs a 2x3x4 factorial design. The factors are `feedback` condition, `pose`, and `duration`. The levels for each of those are vision+proprioception and proprioception $\{Eo, Ec\}$; fully-extended (Straight), pronated vertical bend (Bent), pronated lateral bend (Wing) $\{St, Bt, Wg\}$; and durations of $\{15, 30, 45, 60\}$ seconds. The three levels of the `pose` factor are depicted in Figures 5.3(a)–(c).
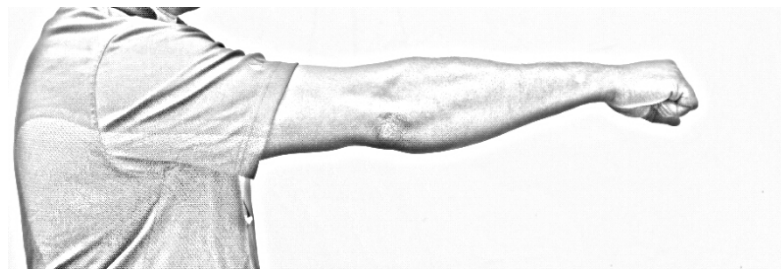
(a)

(b)

(c)

Figure 5.3: Images depicting the three levels of the pose factor variable: (a) fully-extended, (b) pronated vertical bend, and (c) pronated lateral bend.

Consistency in pose configurations across subjects was maintained by specifying that the distance between the wrist and shoulder, for both `Bt` and `Wg`, at onset of the trial was 66% of the distance between the wrist and the shoulder at `St`.

While in a seated position subjects were instructed to raise their arm to shoulder height. The subject's arm was then set to required pose configuration so that the target object was within the field of view of the camera. The height of the chair was adjusted prior to the onset of experimental trials, for each subject, so that the plane containing the dorsal side of the hand and the shoulder was parallel to the transverse plane while the camera was pointed at the target. As each trial commenced, the subject attempted to maintain their original position based on the given `feedback` condition as the reference throughout the allotted duration.

The experiment was partitioned about the `feedback` condition. The first part, employing the `Eo` feedback, consisted of a sequence of 12 `<pose>` x `<duration>` trials which were randomized to prevent any bias in arm exertion related to pose. Between each trial within a part, the subject was given a 30 seconds rest period.

The second part required the subject to perform the same sequence of trials under the `Ec` feedback condition. The same sequence order was used for both parts to ensure that exertion levels were similar between subjects over the course of trials in each part. Parts 1 and 2 of the experiment were conducted on different days to remove cumulative exertion bias from performing all part 1 trials before commencing part 2. The subjects were only allowed to participate in the experiment on days when they did not experience any strenuous arm activity such as an exercise regiment or heavy manual labour, to minimize exertion bias. For each part, the subject maintained a nominal distance of 60cm between the end-effector and the target within the task space to minimize change in pixel pitch.

5.2.3   Results and Discussion

Results were collected from 8 subjects (4 men and 4 women randomly assigned designations of A01 to A08) ranging in age from 25 to 39 years. During each trial, the image-space point of registration, $\mathbf{p}_i = (x_1, x_2)$, was recorded for each frame captured. Three measures were calculated from the image-space position data. These are the instantaneous drift magnitude, $||\mathbf{m}_i||$; the instantaneous drift direction, θ_i; and the cumulative drift, $\hat{\mathbf{p}}$.

$$||\mathbf{m}_i|| = ||\mathbf{p}_{i+1} - \mathbf{p}_i||$$

$$\theta_i = \tan^{-1}\left(\frac{x_{2\,(i+1)} - x_{2\,(i)}}{x_{1\,(i+1)} - x_{1\,(i)}}\right)$$

$$\hat{\mathbf{p}} = \sum_{i=1}^{N-1} \mathbf{m}_i$$

By defining $\mathbf{m}_i$ as the instantaneous drift of the target registration for the i-th frame, we can say that no cumulative drift due to disturbance is observed if $\sum_i^{N-1} ||\mathbf{m}_i|| = 0$. Since the desired control is to maintain a fixed target registration which is applied at each frame, we make the assumption that $\mathbf{p}_{i+2}$ and $\mathbf{p}_i$ would be uncorrelated in that the movement given by $\mathbf{m}_{i+1}$ is not a necessarily a co-activation compensating for the movement $\mathbf{m}_i$.

Figure 5.4 gives an example of the set of graphs for one trial performed subject A01. A compass plot giving both magnitude and direction of the instantaneous drift vectors is provided in Figure 5.4(a). Since the multiple instances of equivalent instantaneous drift vectors are plotted over top of each other, histograms of the corresponding magnitude, $||\mathbf{m}_i||$, and direction, θ_i, are provided in Figures 5.4(b)–(c), respectively. It is important to note that the direction is calculated based on the image-space coordinates of a target which is stationary in task-space so the actual task-space camera movement is the negative of $\mathbf{m}_i$. Thus an angle of $\theta_i = 0$ represents a horizontal movement of the camera to the medial line of the body (leftward) for right-hand use.

Experiment trials data were partitioned by `<pose>` x `<feedback>` combinations and the mean of the instantaneous drift vector magnitudes was calculated for for all subjects within each `<duration>`. One-way ANOVA was performed using Matlab's Statistical Toolbox. The results are provided in Table 5.3, and show that there is no statistical difference between the mean magnitude of instantaneous drift across subjects maintaining a statically loaded pointing task (target registration) within each of the six `pose` x `feedback` factor combinations. This indicates that there is no significant effect on target registration stability due to exertion up to 60 second intervals of static loading.

The cumulative drift vector data, partitioned by `<feedback>` condition, is shown Figure 5.5. Examination of the angular histogram plots in Figures 5.5(b) and 5.5(d) shows a similar lateral drift bias towards the body mid-line and similar vertical drift distribution. This indicates that the inclusion of an end-point close-loop visual reference of has no significant effect on cumulative drift direction under static load. However, the change in `<feedback>` condition exhibits a significant effect

Table 5.3: One-way ANOVA results of mean instantaneous drift magnitude during static loading, partitioned in `pose x feedback` combinations.

Partition	15sec	30sec	45sec	60sec
`<Bt>x<Eo>`	1.3344	1.3967	1.3214	1.2886
	$F(3,28)=0.28$, $p=0.8373$, $se=0.0851$			
`<St>x<Eo>`	1.6609	1.6548	1.39	1.5228
	$F(3,28)=0.91$, $p=0.447$, $se=0.1343$			
`<Wg>x<Eo>`	1.6015	1.3063	1.4491	1.3254
	$F(3,28)=0.7$, $p=0.5582$, $se=0.1624$			
`<Bt>x<Ec>`	1.4932	1.3351	1.1992	1.2414
	$F(3,28)=2.3795$, $p=0.0909$, $se=0.0845$			
`<St>x<Ec>`	1.3201	1.4819	1.4703	1.4548
	$F(3,28)=0.51$, $p=0.6768$, $se=0.1051$			
`<Wg>x<Ec>`	1.2596	1.2179	1.1352	1.1045
	$F(3,28)=0.61$, $p=0.6122$, $se=0.0918$			

on cumulative drift magnitude as seen across Figures 5.5(a) and 5.5(c) with mean magnitudes of 45 and 27 pixels, respectively. This indicates a weaker ground truth reference accuracy with proprioceptive feedback only.

(a)

(b)

(c)

(d)
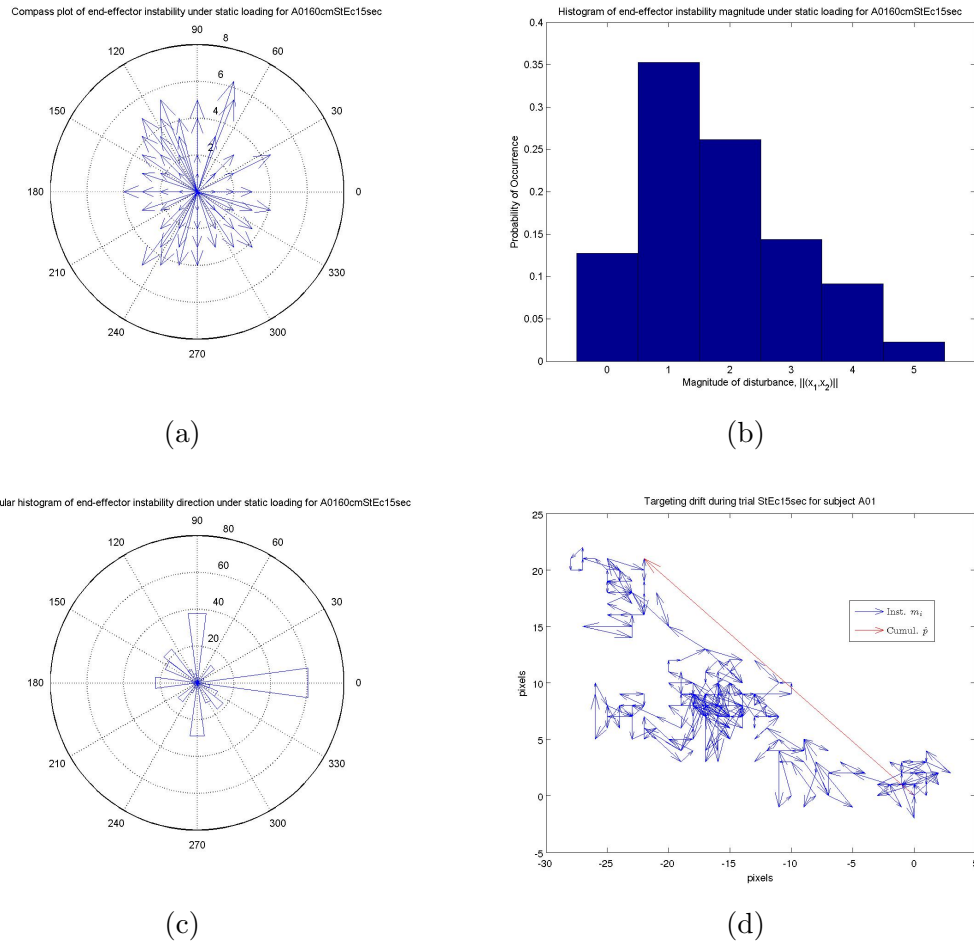
Figure 5.4: Graphical depiction of the experimental trial performed by subject-A01 under the conditions of <St>x<Ec>x<15sec>: (a) compass plot of instantaneous drift vectors $\mathbf{m}_i$, (b) histogram of instantaneous drift vector magnitudes, (c) angular histogram of instantaneous drift vector directions θ_i, and (d) showing the individual drift vectors end to end and the cumulative drift $\hat{\mathbf{p}}_i$.
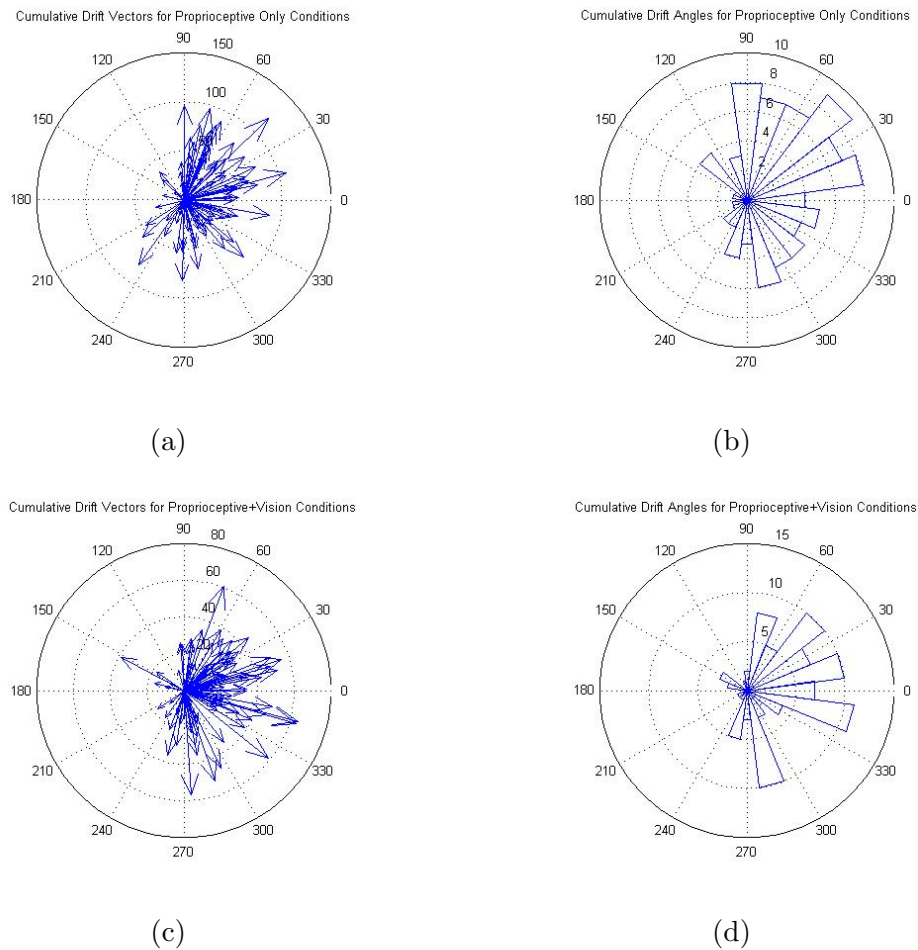
Figure 5.5: Graphical depiction of the experimental trials performed by subjects A01 through A08 partitioned by <feedback> condition: (a) compass plot of cumulative drift vectors for all proprioceptive only conditions, (b) angular histogram of cumulative drift vector directions for all proprioceptive only conditions, (c) compass plot of cumulative drift vectors for all proprioceptive+vision conditions, and (d) angular histogram of cumulative drift vector directions for all proprioceptive+vision conditions.

5.3 Experiment 3: DES-Controller with tactile frequency variation

A revised prototype was constructed to expand upon the experimental capabilities of the system. The revisions were based on qualitative feedback and gathered from the results of testing the initial version described in Section 5.1. Revisions to the both the hardware and software implementations were made. The custom written computer vision application was replaced with a commercial package, RoboRealm[5]. The package has an extensive library of filters which can be layered in sequence within an imaging processing pipeline to perform a variety of comparable filtering, feature extraction, and object detection tasks at higher frames than the initial prototype. This update was necessary to minimize the frame processing time, resulting in a minimal lag when issuing motion cues to the user. A custom control application implementing the both the l_1- and l_2-models of the DES-Control scheme proposed in Chapter 3 was written using Microsoft Visual Studio 2008 (C++).

The RoboRealm application executes a series of filter objects as an image processing pipeline to perform image acquisition, colour segmentation, noise removal, object identification, and feature extraction for each frame captured. The pipeline is executed upon request from the control application via a TCP/IP socket connection. The extracted features correspond to feature vector, $\mathbf{f}_i = \begin{bmatrix} p_x \ p_y \ f_z \end{bmatrix}^T$, as described in Section 2.4.

The initial prototype was only capable of implementing bang-bang control due to a limitation of the SBC. The revised prototype uses a Freescale HCS12 based microcontroller board in the wearable portion of the system. The DP256 model of the HSC12 microcontroller includes a PWM peripheral with a sufficient number of channels to support implementing a proportional control input signal for each vibrotactor.

The images in Figures 5.6(a) and 5.6(b) show the wearable portion and the MS-Windows based control software portion of the second iteration of the aiReach system prototype, respectively. The screen capture image in Figure 5.6(b) shows the two software applications that make up the computer vision pipeline and DES-

---

5. available at www.roborealm.com

Controller. This prototype version was used for the remainder of the control scheme experiments presented in this chapter.

(a)   (b)
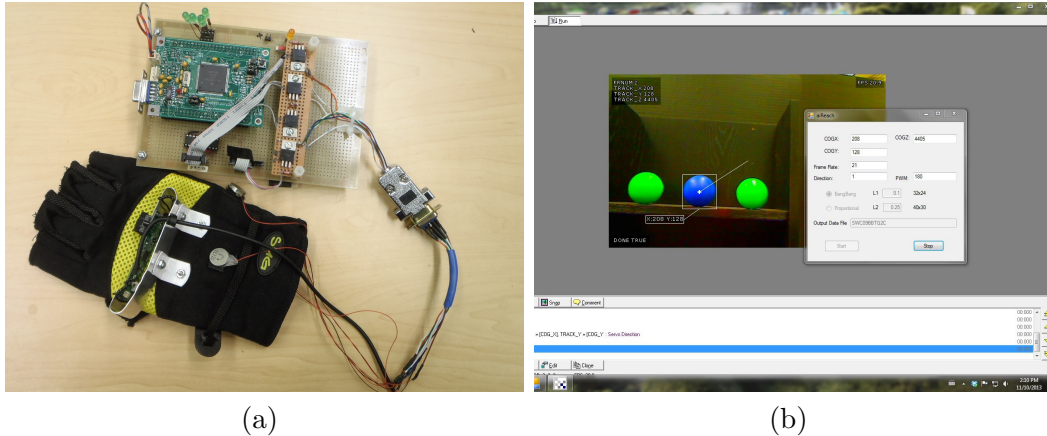
Figure 5.6: Images depicting the hardware and software components of the second iteration of the aiReach prototype: (a) the wearable component including the HC12S based microcontroller board, glove, camera, and vibrotactors; (b) screen capture of the MS-Windows based software including RoboRealm vision platform (background) and custom controller application (front-right).

Since the feasibility of the aiReach system, as a wearable guidance aid for reaching tasks, was proven in the initial prototype testing, the purpose of the testing of the revised prototype is to determine the difference, if any, in the learned response of the user to the bang-bang versus proportional control feedback scheme. The experiment investigates whether the semantic representation of distance through the use of varying the frequency of vibration in proportion to the targeting tracking error and movement time is a more responsive control scheme.

5.3.1   Experimental Apparatus

The apparatus used in this experiment is depicted in Figure 5.7. The rig is a two level shelving unit 38 x 46 x 17cm, with 21cm separation between the shelves. The possible mounting positions for the target are 10cm apart along each shelf to allow for differing angles of approach. The rig allows for six possible target positions (TG1 - TG6) as numbered in Figure 5.7 for placing the spherical targets. The rig is mounted on a tripod stand to allow height adjust so that the midpoint of the back wall can be vertically aligned with the subject's shoulder height. Figure 5.7 also shows labels

for projections of the four vertically and horizontally aligned start positions (A - F) corresponding to a target at TG1.
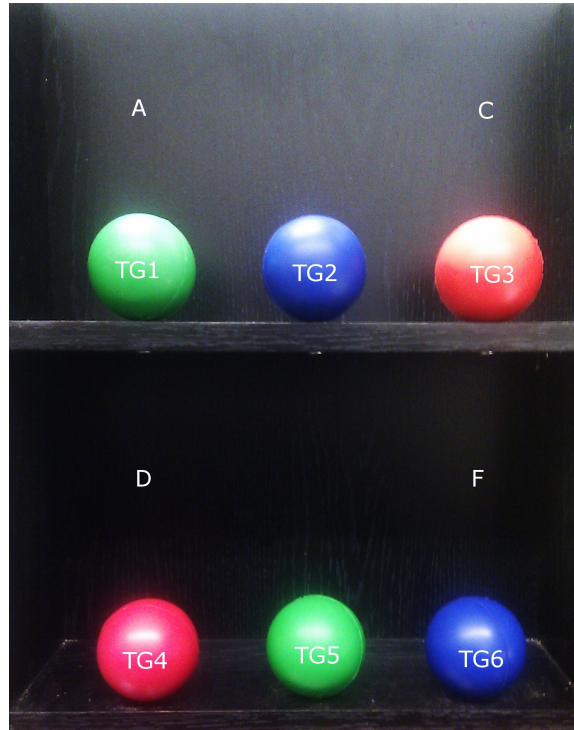
Figure 5.7: The experimental apparatus consisting of a two level shelf with an six possible positions for target placement.

5.3.2 Experimental Procedure

The experiment employs a 2x3x4 factorial design. The factors are `signal` condition, `target` position, and `start` position. The levels for each of those are bang-bang (`BB`) and proportional (`PR`) vibrotactile signalling; three of six, randomly selected target positions on the rig; and four start positions corresponding to the selected target position. The colour segmentation based tracking technique described in Section 3.4 was used to achieve a sufficiently high frame rate.

Prior to the onset of the reaching task the subject's vision is blocked and the target (blue) object is placed at one of the randomly selected `TG` positions on the test rig. At least two other spheres of different colours are also randomly placed at another available `TG` position(s). The subject's hand is placed at an initial position

within the task space that vertically and horizontally aligns with one of the adjacent ST positions. For each of the $k = \{1 \ldots 24\}, ||\mathbf{p}_k|| = L$, away from the target. The cartesian distance L remains constant within the set of trials. By randomizing $\mathbf{p}_k = [x \; y \; z]$ such that $||\mathbf{p}_k|| = L$, the axial component distances are randomized; avoiding a biasing of performance results through the subject's kinaesthetic memory acquired from performing the same pattern of movements, repeatedly. As an example, the set of starting position alignments for TG2 are shown in Figure 5.7. This spacial relationship between start and target position requires that three axial displacements (horizontal, vertical, and depth) be resolved to complete the reaching task trial.

The DES-control application shown in Figure 5.6(b) can be configured for operation under either a l_1- or l_2-model, with either bang-bang or proportional CT-plant input signalling. For the purpose of this experiment, the configuration is fixed to operate under the l_1-model with ($l_1 = 0.1$, $l_2 = 0$).

At the onset of a trial, tracking data for each video frame is logged by the control-tracking software. Tracking data includes frame time-stamps, the feature space vector, hypersurface locations, and CT-plant control input. The data is recorded until the reaching task is completed. The reaching task is considered complete under two possible circumstances: either the subject's hand comes into contact with the target or the STOP motion cue is issued. All trials recorded were of completed reaching tasks. If the subject did not adequately complete the reaching task the trial was repeated.

Subjects are given three practice trials before starting the set of 24 recorded trials. The first practice trial is performed sighted so the subject can correlated spacial awareness of their hand movements to their perception of the motion cues. The remaining two practice trials are performed unsighted. Subjects are given a minimum of 30 second rest between trials. Subjects are given a longer rest period between trials if desired.

5.3.3 Results and Discussion

Data was collected from 9 subjects: 6 males and 3 females (23 - 39 years old). No subjects requested additional rest time. Each subject performed the set of 24 trials, 12 under bang-bang and 12 under proportional signalling condition. Five subjects

(B03,B06,B07,B08, and B09) began with the block of 12 trials under the bang-bang signalling and the other four (B01,B02,B04, and B05) began with proportional block of trials. In all cases, subjects were informed which signalling method was being used prior to commencement of each block of 12 trials. Of the total 216 trials recorded across all subjects, only 12 were repeated attempts.

A significant limitation with the initial prototype regarding tracking performance was the frame rate for the entire control loop. With the implementation of the second prototype, the average frame rate for feature extraction within RoboRealm achieved 30 fps. Incorporating the the additional execution of the DES-Controller application and communication with the microcontroller unit, the average frame rate for the entire control loop was 22 fps.

5.3.3.1   Qualitative Observations

A number of qualitative observations regarding the subjects' response to the interface and performance made during the experimental sessions are important to note. These observations aid in the interpretations of the graphical results provided in the following sections. Amongst the nine subjects: six used a straight-arm posture during the trials, while two used a slightly bent posture, and one used a wing posture at commencement of each trial. Reaching posture variation was entirely up to the choice of the subject. In all cases, the reaching task distance of 60cm required the subjects to take at least one step forward to complete the task, with the norm being two or more steps. Almost all subjects exhibited moderate to significant timidity in moving forward during the depth tracking phase(s) of the reaching. It was demonstrated through the behaviour of attempting to lean forward by bending at the waist with the stance fixed in place. This could be seen to degrade steady arm movement capability as the subject became moderately unbalanced due to a shift in center of gravity. The subsequent forward step from the leaning position would result in a bounce in hand (camera) motion. This type of behaviour can be see in the trajectory tracking plot shown in figure 5.8(a). An alternate forward movement behaviour of one large step from the initial stance also manifested a similar bounce in the camera motion.

Another, more interesting anomaly found in the trajectory tracking plots is shown in figure 5.8(b). The behaviour is a strong lateral ulnar deviation angle in the wrist posture shown in figure 5.8(c). This strongly manifests itself in the latter
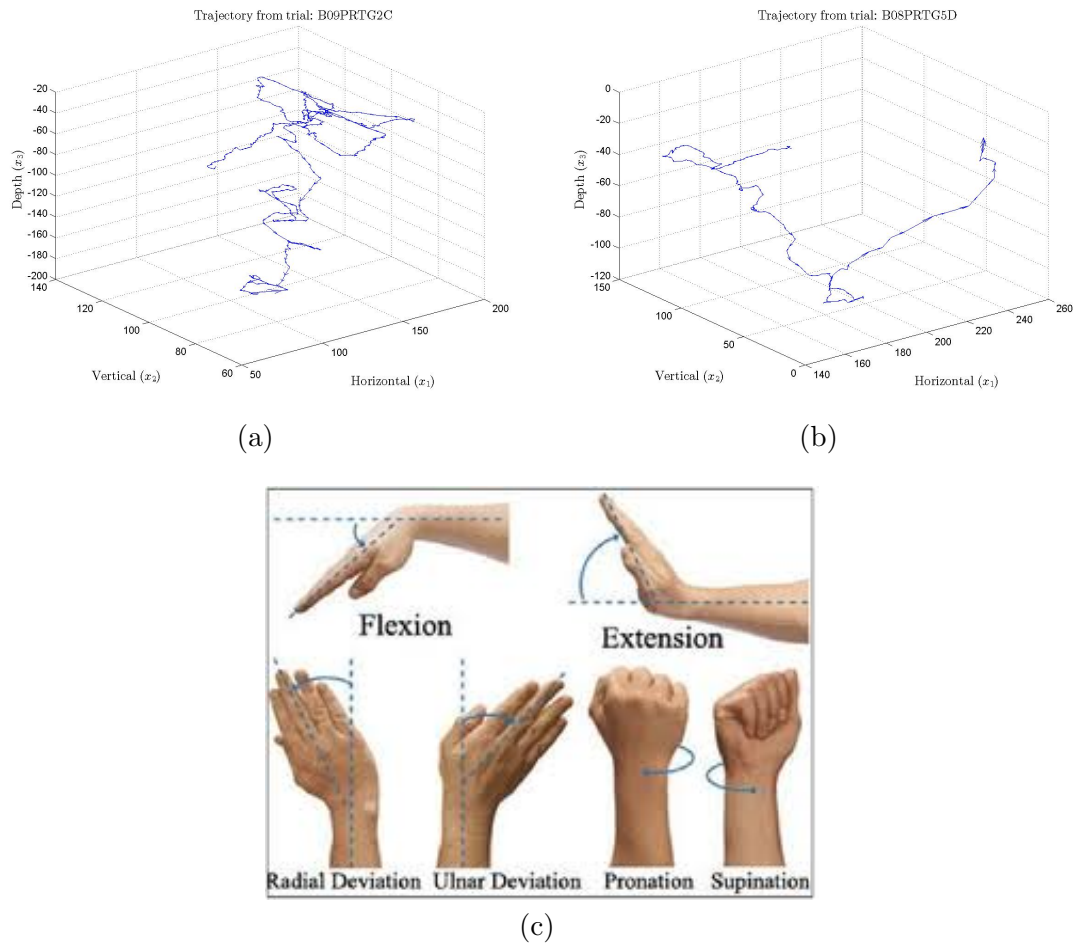
Figure 5.8: Samples of typical anomalous depth tracking behaviours in trajectory plots: (a) abrupt bounce due to forward step; (b) false backward movement due to ulnar deviation; (c) anatomical wrist postures

portion of the depth tracking phase by directing the optical axis, and thus the field of view of the camera, away from the normal of the egocentric frontal plane.

As such, a greater and greater portion of the segmented target within the image is occluded off-screen. Since the feature extraction technique calculates f_z, the depth tracking feature, as the diameter of the segmented pixel mass, the value of $x_3 = -f_z$ will actually increase[6] even as the camera proceeds towards the target. This can be seen in the trajectory tracking sample plot provided in figure 5.8(b). The plotted tra-

_____

6. Decreasing values of x_3 indicate motion toward the target to maintain a right-handed coordinate frame into the image

jectory seems to indicate that the forward motion actually reversed and the terminal point of the reaching task is farther away from the target than the starting position of the task. However this is known to not be the case as all trials performed during the experiment terminated at the target. Of the 216 trials, 39.8% exhibited this type of false-backward behaviour in the trajectory plots.

Some vertical flexion and extension deviations were observed as well, but in those instances the subjects were observed to more readily perceive and correct the inclination angle. The qualitative observations of wrist posture deviations indicates that the wrist posture instability is a very compelling issue in the control of an eye-in-hand configuration. However, the distributed contribution of vertical and lateral deviations in wrist posture stability may not be readily controllable by the subject with training and/or augmentation of the wearable components of the system. Further study must be undertaken to develop and test mechanisms for stabilizing wrist posture during the reaching task.

5.3.3.2   Trajectory and Performance Analysis
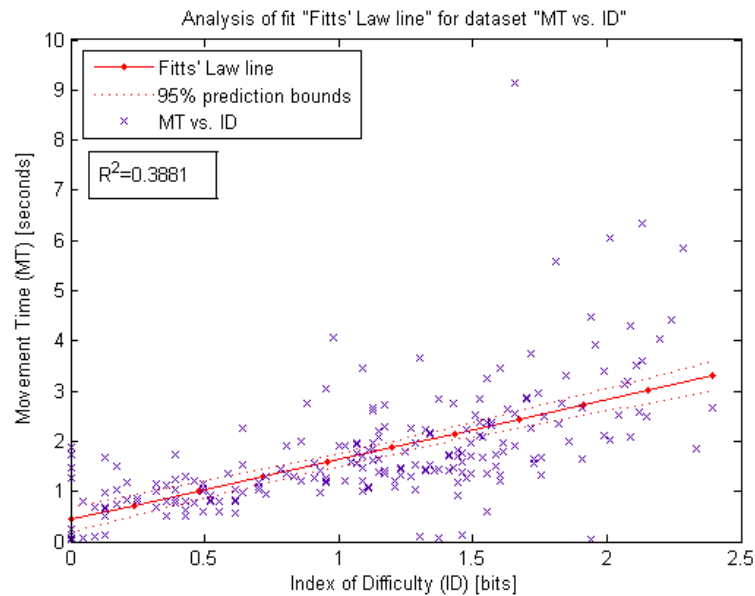
Figure 5.9: Scatter plot of movement time (MT) versus index of difficulty (ID) for the initial horizontally or vertically resolved tracking movements from each trial. The plot includes the Fitts' Law linear approximation and 95% confidence bound

For each trial, the ID was calculated for the initial DES-controller state change resulting in the resolution of a 1-dimensional reaching task. Figure 5.9 shows a plot of the ID for those 1-dimensional reaching task versus their corresponding movement time, MT. The plot also includes a regression line with 95% confidence bounds which approximates the 1-D Fitts' law performance model. The linear regression shows a very poor fit to the data with a $R^2 = 0.3881$. This is well below the typical fit ($R^2 \approx 0.9$) level common in most other Fitts' law performance studies [110, 118, 112].
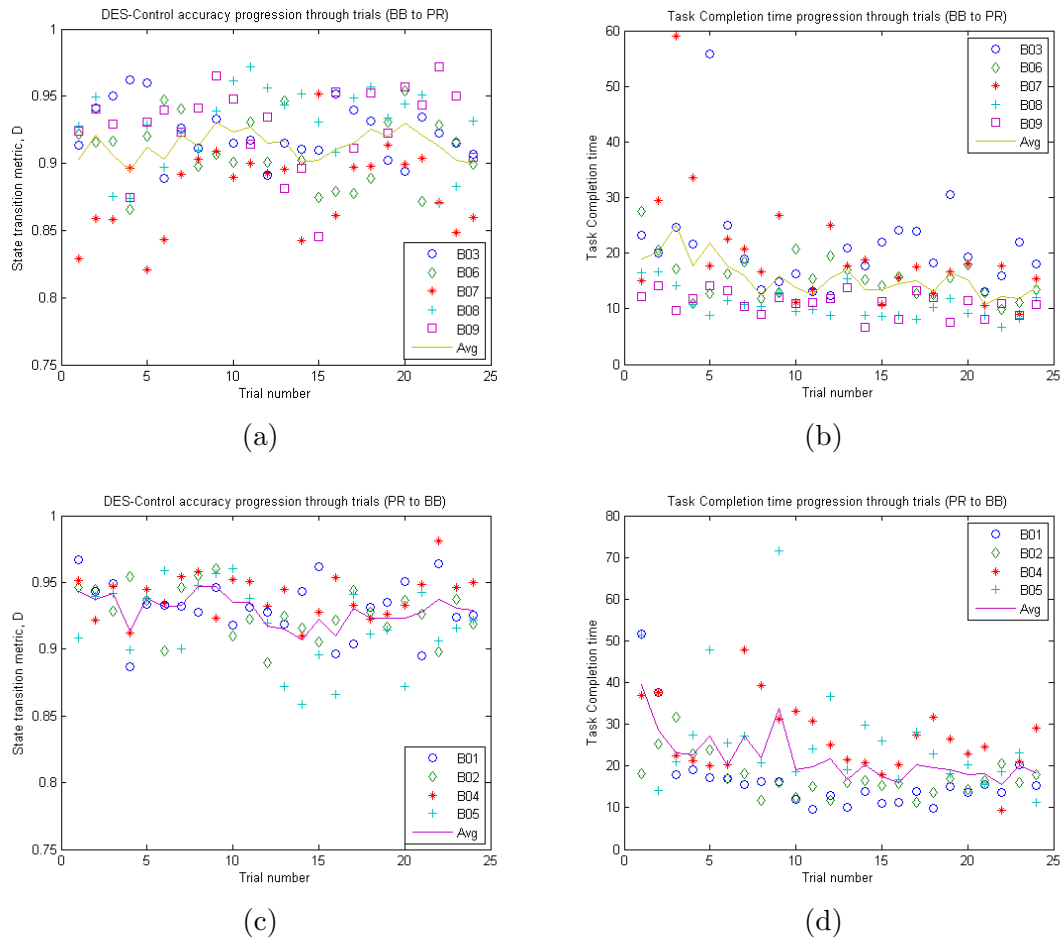
Figure 5.10: DES-Controller state transition accuracy metric and completion time progression across the sequence of trials: (a) Accuracy for bang-bang to proportional progression; (b) Completion time for bang-bang to proportional progression; (c)) Accuracy for proportional to bang-bang progression; (d) Completion time for proportional to bang-bang progression

The plots in figure 5.10 partition the reaching task trajectory data by subject

groupings of those that started the session with the bang-bang (BB-PR) signalling condition from those that started with the proportional (PR-BB) signalling condition. The purpose is to determine if there is any discernable effect on performance due to signalling condition. The implication being, whether perceptual coding of tracking distance through vibration frequency significantly affects reaching task performance.

Figures 5.10(a) and 5.10(b) give the calculated state transition metric, D, (see equation 3.56) and the task completion time for the BB-PR grouping by trial number. Similarly, figures 5.10(c) and 5.10(d) display the equivalent data for the PR-BB grouping of subjects. In both groupings we see a general decreasing trend in task completion time as subjects perform a greater number of trials, but a generalized constant trend in state transition metric over the same span of trials. This seems to indicate that the type of haptic signalling did not have a significant effect on average subject performance. Qualitative comments provided by the subjects during the experiment indicate that in many instances subjects were not consciously aware of differences in vibration frequency. For a majority of subjects, their primary concentration was focused on discerning which directional cue was issued, and little or no attention was paid to the frequency of vibration of the cue. However, inspection of the averaged line of both figures 5.10(a) and 5.10(c) near the twelfth trial shows an interesting result. It is the point at which the signalling type changed for each grouping of subjects. In both graphs there is drop in the averaged state transition metric (accuracy) score after the signalling type change and then a gradual increase over the remaining trials. This could be attributed to an unconscious perception of the signalling type and resulting necessity to adapt to the new coding scheme for tracking distance. This implies that subjects' tracking accuracy performance decreased with the introduction of a different signalling type, but increase again as the subjects became familiar with it. This indicates that while subjects were not always conscious of the difference between the two displacement magnitude encoding schemes, it was perceived on some subconscious level. The tracking performance as measured by average task completion time does see improvement due to cumulative practice (across 24 trials). The result is to be expected as performance should increase with successive practice of a motor task.

The plot in figure 5.11(a) shows the task completion time versus the state transition metric for each of the 216 trials. The plot demonstrates a speed versus accuracy relationship analogous to Fitts' law with inclusion of a regression line that shows a
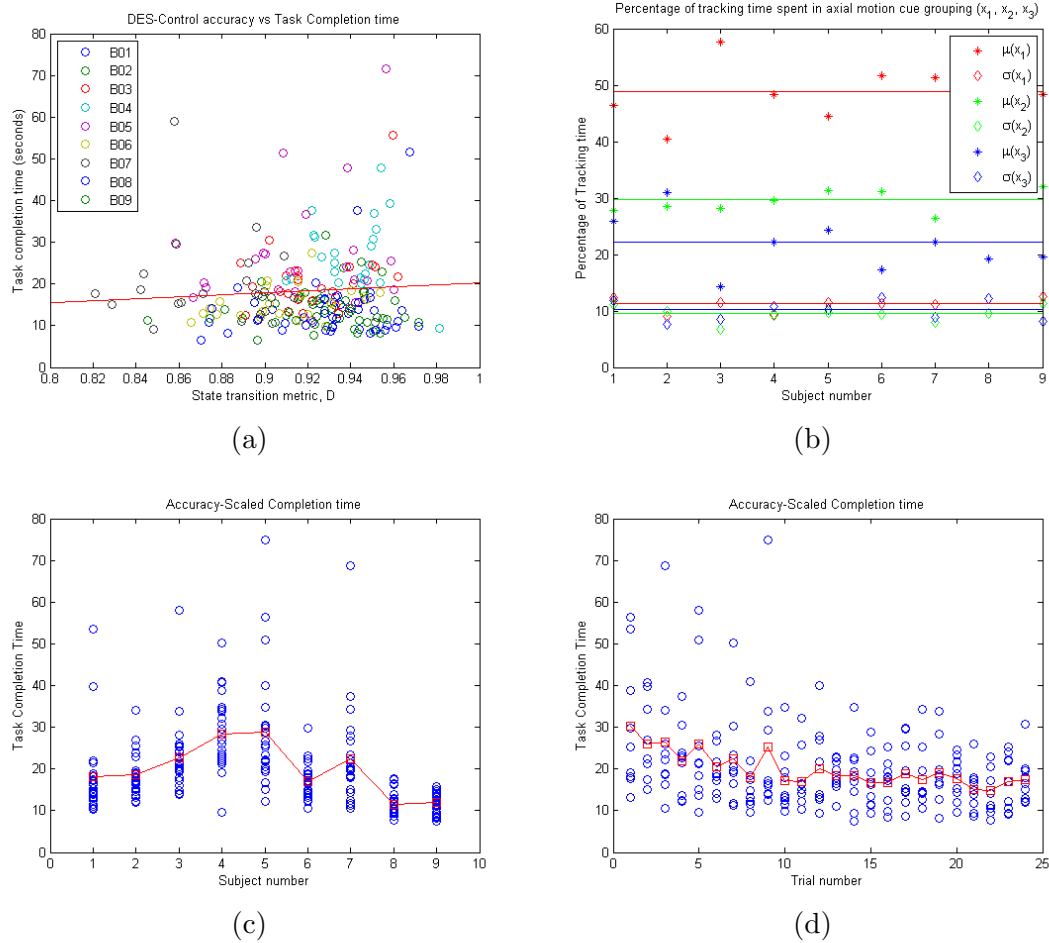
Figure 5.11: Completion time graphs: (a) State transition metric vs Measured task completion time; (b) Tracking time statistics by axis; (c) Scaled completion time by subject; and (d) Scaled completion time by trial.

generalized relationship trend of increased completion time as the accuracy measure of tracking increases. However, the speed-accuracy relationship demonstrated here shows that it characterizes a task that is different from a classical ballistic reach to a target. With a Fitts' Law reaching task the terminal point location is known and at the target object. With the reaching task performed here, neither the target object location nor the terminal point of the full task or any subtask is known. Essentially the task goal is to track a variable length sequence of directional cues. Each non-ideal tracking of a motion cue lengthens the sequence by at least one additional DES-controller state transition and a variable length of time. As a consequence, trajectories with the identical state transition metric scores can have vastly different

completion times which is somewhat contradictory to monotonic relationship embodied in Fitts' law. Given that completion time for a reaching task is highly variable due to many factors, not the least of which is cumulative learned response, we present the scaled completion time in figures 5.11(c) ordered by subject, and 5.11(d) ordered by trial number. The accuracy-scaled completion time for each trial is calculated by dividing the measured completion time by it's corresponding state transition score. The ordering by subject in figure 5.11(c) shows significant performance differences that can occur within difference plants (users) while preforming the same set of guided reaching tasks, but by comparison to figure 5.11(d) we see the a strong convergence in this task accuracy measure due learned response.

The plot in figure 5.11(b) shows the mean and standard deviation of percentage tracking-time per movement axis across all 24 trials for each of the subjects. The Root-Mean-Squared (RMS) values for mean (dashed-line) and standard deviation (solid-line) are included in the plot. The large difference in RMS mean values in conjunction with the very similar RMS standard deviation values indicates a bias in the percent tracking-time across the three axes. This is to be expected given that the hybrid-control scheme is prioritized to correct tracking error in $\mathbf{x_1}$, then $\mathbf{x_2}$, then $\mathbf{x_3}$. However, taking into the account the stronger mid-line ($\mathbf{x_1}$) bias in the static loading experimental results presented in Section 5.2, it is likely that the bias in lateral correction is a combination of the prioritized hybrid-control scheme and natural proprioceptive factors. It is not possible to determine what contribution each factor makes towards this bias from the data available in this experiment. Doing so would require implementing an alternate experiment configuration that would generate results indicating whether or not an inherent natural stability bias exists, and to what extent, between the three axial motion cue groupings; which we discussed further in chapter **??**.

The results provided above demonstrate the measure of accuracy with which the plant (user) can respond to DES-state transitions between the various one dimensional tracking controllers, but not the precision of the tracking response within a particular state, i.e. how precisely the camera motion follows a lateral path while in $\tilde{s}_2$. To describe this plant behaviour we must examine the directional error fraction (DEF), given by equation 3.58, of all trajectory segments for the corresponding DES-Controller output symbol as an input to the actuator function, $\gamma(\tilde{r})$ (see equation 3.19). Figure 5.12 gives a sample of the set of 5 DEF plots for the single tracking

trial B02BBTG2A as an example. There are a number of important aspects of the DEF plots to consider for interpretation:

- the length of the horizontal axis of each plot corresponds to the number of trajectory segments for which the system was operating under a particular DES-controller output symbol;

- the scalar quantity is plotted for each of the motion cue axis per trajectory segment;

- the sum of the 5 plot lengths is equal to the total number of trajectory segments for that trial; and

- one or more of the plots can be empty, indicating that that particular DES-controller output symbol was never (or infrequently) generated.

For ideal behaviour, the magnitude of $\rho_{x_i}[n]$ should be one with the sign dependent on the intended direction of motion given by the DES-controller output symbol, but for all other $\rho_{x_j}[n] = 0, j \neq i$. Thus all movement during the n-th trajectory segment was directed precisely in alignment with the motion cue issued by the controller.

The descriptions given in Table 5.4 provides the basic interpretation from the egocentric reference frame for the value of each DEF per DES-controller output symbol. To maintain a right-hand coordinate frame, procession towards the target is given by a decreasing negative value along the x_3 axis. Therefore, for the output symbol $\tilde{r}_1$ which gives the `Move-Forward` motion cue, the tuple of $\{\rho_{x_1}, \rho_{x_2}, \rho_{x_3}\} = \{0, 0, -1\}$ represents ideally-behaved tracking of the intended trajectory.

By applying the appropriate sign change to the ρ-tuple for trajectory segments associated with $\tilde{r}_1$, $\tilde{r}_3$, and $\tilde{r}_5$ output symbols in the trajectory we can consistently attribute well/ideal behaviour to a positive $\rho_x[n]$ value along the intended motion cue axis. This allows us to generate the average precision metric across all the trajectory segments within a task trial, and thus compare it to other task trials regardless of their length. To do so we plot the mean DEF ($-1 \leq \theta_e \leq 1$) for each trial.

Figures 5.13(a) and 5.13(b) give the mean DEF plotted by subject and by trial progression, respectively. Each point within the plot gives the averaged degree of precision of motion of the end-effector as driven by the CT-plant (internal joint-space

DES Output symbol	Dir. Err. Frac.	target offset from image center
$\tilde{r}_1$	$\rho_{x_1} > 0$	right
	$\rho_{x_1} < 0$	left
	$\rho_{x_2} > 0$	above
	$\rho_{x_2} < 0$	below
	$\rho_{x_3} > 0$	ill-behaved
	$\rho_{x_3} < 0$	well-behaved
$\tilde{r}_2$	$\rho_{x_1} > 0$	well-behaved
	$\rho_{x_1} < 0$	ill-behaved
	$\rho_{x_2} > 0$	above
	$\rho_{x_2} < 0$	below
	$\rho_{x_3} > 0$	away
	$\rho_{x_3} < 0$	closer
$\tilde{r}_3$	$\rho_{x_1} > 0$	ill-behaved
	$\rho_{x_1} < 0$	well-behaved
	$\rho_{x_2} > 0$	above
	$\rho_{x_2} < 0$	below
	$\rho_{x_3} > 0$	away
	$\rho_{x_3} < 0$	closer
$\tilde{r}_4$	$\rho_{x_1} > 0$	right
	$\rho_{x_1} < 0$	left
	$\rho_{x_2} > 0$	well-behaved
	$\rho_{x_2} < 0$	ill-behaved
	$\rho_{x_3} > 0$	away
	$\rho_{x_3} < 0$	closer
$\tilde{r}_5$	$\rho_{x_1} > 0$	right
	$\rho_{x_1} < 0$	left
	$\rho_{x_2} > 0$	ill-behaved
	$\rho_{x_2} < 0$	well-behaved
	$\rho_{x_3} > 0$	away
	$\rho_{x_3} < 0$	closer

Table 5.4: Egocentric reference for directional error fraction plots per
DES-controller output symbols

controller). If the end-effector were to perfectly follow the motion cues issues through DES-controller outputs the mean DEF would be $\theta_e = 1$. Conversely, if the exact inverse motion for every motion cue were performed the mean DEF for that particular trial would be $\theta_e = -1$. However, the latter example is not practically feasible in that the trial would unsuccessfully terminate with loss of the target from the field of view. A value of $\theta_e = 0$ indicates that the CT-plant's directional precision is neutral, which means that on average within that reaching task, the user moved along the intended trajectory with positive correspondence equally as much as they did with negative correspondence to the motion cues. The measure of precision is purely directional (relative position) as CT-plant is not capable of accepting control inputs for absolute position, velocity, or acceleration set-points. The low range of values plotted seem to indicate a very poor, error prone system, but that is only in the context of comparison between a human and an actual robotic manipulator performing the same task. There are a number of contributing factors to the low mean value of the precision metric. The first, but only marginal factor is the natural kinesthetic limitation of the user to perceive how well they are tracking intended motion cues exactly. This factor is shown by the points that fall in the range of $0 < \rho_x[n] < 1$ in the full tracking sequence plots in figure 5.12. A strong contribution comes from the points that fall in the range of $-1 \leq \rho_x[n] < 0$. This factor is manifest of the false-backwards behaviour described previously in Section 5.3.3.1. The strongest contributing factor is from points in the sequence where $\rho_x[n] = 0$, which are manifest of the CT-plant performing no directed motion when cued to do so. These instances are manifest of pauses in arm motion as the user attempts to perceive and decipher the motion cue issued. This inaction in the presence of motion cue is consequently a directional error and accordingly diminishes the measured precision during tracking.

Combining the state transition metric of accuracy and the DEF metric of precision, we can generate a per segment tracking response metric ($-1 \leq \delta[n] \cdot \rho_x[n] \leq 1$) that provides a measure of how well the plant (user) has followed the control (motion cues) per trajectory segment. There are three atypical extremes in tracking performance present that can be described from the plots that follow. The nominal task trial tracking response metric is similarly the mean value of across the particular trial, ($-1 \leq \psi \leq 1$). Figures 5.13(c) and 5.13(d) give plots of the mean tracking response metric by subject and by trial progression, respectively. Figure 5.13(e) gives the mean tracking response vs task completion time for all trials.
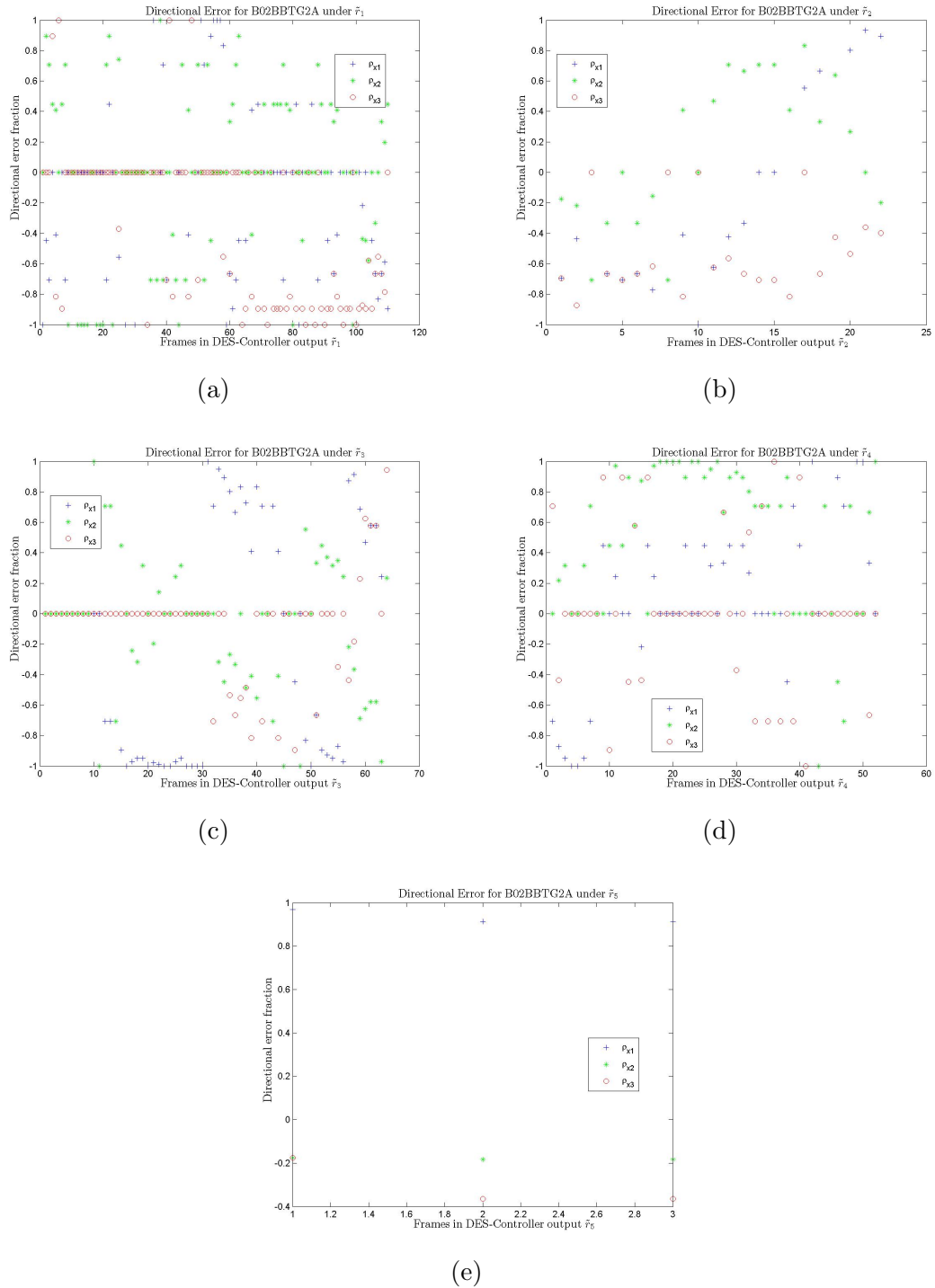
Figure 5.12: Sample trajectory data plots from one trial with subject B02: DEF plots for each of (a) DES-controller output symbol $\tilde{r}_1$; (b) DES-controller output symbol $\tilde{r}_2$; (c) DES-controller output symbol $\tilde{r}_3$; (d) DES-controller output symbol $\tilde{r}_4$; and (e) DES-controller output symbol $\tilde{r}_5$
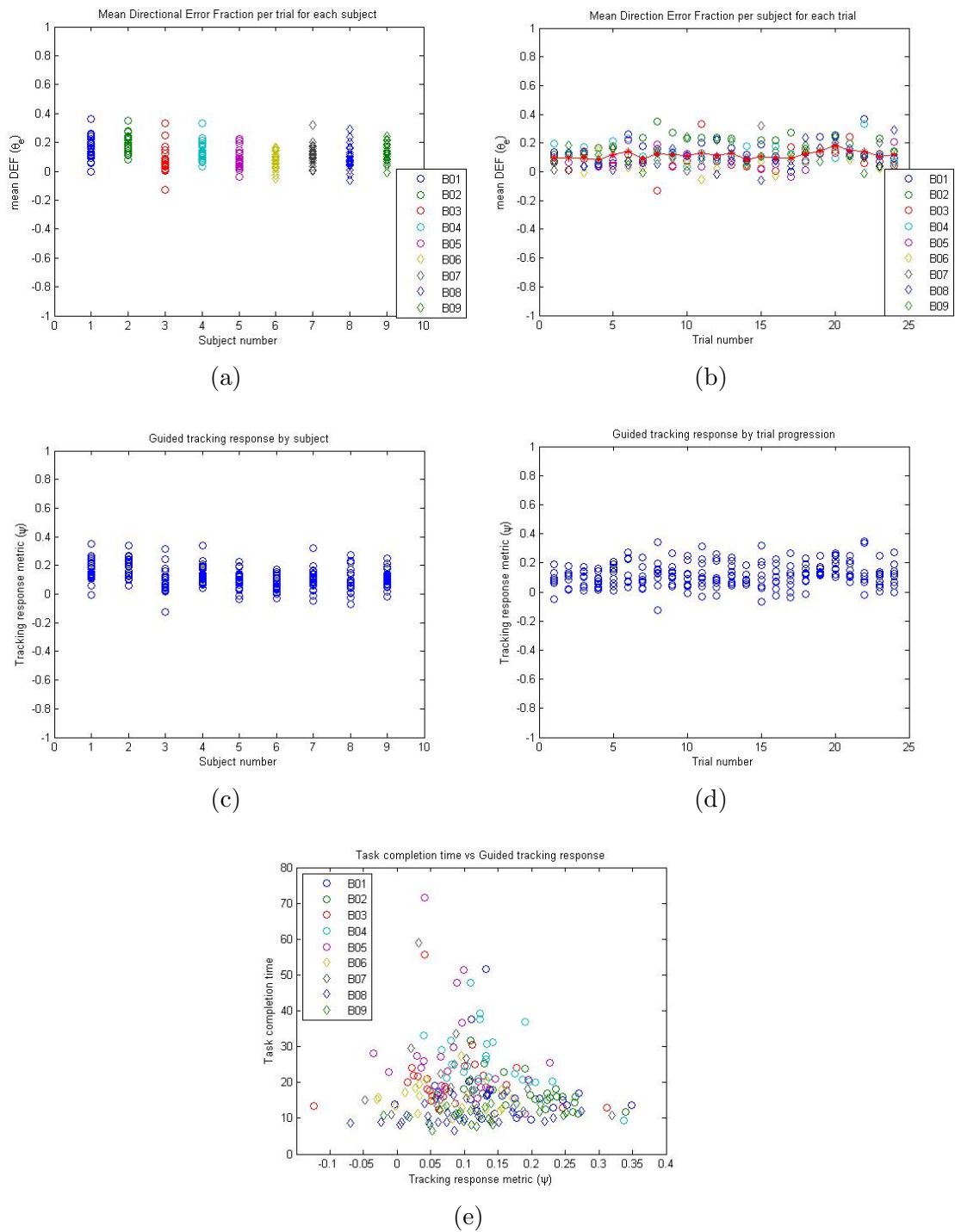
Figure 5.13: Mean Directional Error Fraction (θ_e): (a) by subject, and (b). The mean tracking response (ψ): (c) by subject, (d) by trial, and (e) versus task completion time.

5.4 Experiment 4: DES-controller with l_1 and l_2 parameter variation

In this experiment trajectories from a set of reaching task trials with the spherical targets and shelf rig are analyzed with the proposed discrete event measures proposed in section 3.7 to determine the effect of varying the hypersurface boundary locations within the image state-space. An approximate target hysteresis region is created using non-zero values for the l_2-model parameter. We compare the results from this set of trials to those collected in the previous experiment to verify our hypothesis that the introduction of the hysteresis region allows the user to achieve better guided tracking response levels by mitigating the errant or oscillatory targeting corrections that occur a the target region boundary in the l_1-model configuration.

5.4.1 Experimental Apparatus

The physical apparatus used in this experiment is the same as described in section 5.3.1. However, in this experiment the l_1 and l_2 DES-Controller parameters are varied across 4 sets of non-zero values. The guided reaching task trajectory logs from each trial were captured and their analysis is presented in following sections.

5.4.2 Experimental Procedure

The experimental procedure is very similar to that described in section 5.3.2, but with changes to the set of DES-Controller parameters.

The experiment employs a 2x4 block design. Target and start position pairings are randomly selected for each trial, maintaining a $||\mathbf{p}_k|| = 60$cm as the initial straight line magnitude of the reaching task. The same sequence of start and target positions specified in the previous experiment were used here. The factors of `signal` condition, and DES-controller parameters: l_1, and l_2 were randomly assigned to each trial. The `signal` condition varied between bang-bang and proportional so that 12 trials under each were performed by each subject during their session. The DES-Controller parameters of l_1 and l_2 were varied amongst 4 sets: $(0.075, 0.75)$, $(0.1, 0.25)$, $(0.1, 0.5)$, and $(0.125, 0.2)$, so that 6 trials under each were performed by each subject. The hypersurface locations corresponding to the parameter values are given in Table 5.5.

Table 5.5: Values for l_1 and l_2 parameters and corresponding hypersurface locations used in Experiment 4

Target Window (l_1, l_2)	$h_2(x)$	$h_1(x)$	$h_3(x)$	$h_4(x)$	$h_6(x)$	$h_5(x)$	$h_7(x)$	$h_8(x)$
$(0.1, 0.25)$	140	144	176	180	105	108	132	135
$(0.1, 0.5)$	136	144	176	184	102	108	132	138
$(0.075, 0.75)$	139	148	172	180	104	111	129	135
$(0.125, 0.2)$	136	140	180	184	102	105	135	138

Consistent with the previous experiment, subjects were given 3 practice trials before formal trials began. The first practice trial under a sighted condition, and the other two unsighted. Assignment of the DES-Controller parameter set, signalling condition, as well as start and target position pairs were randomized during the three practice trials to avoid biasing the subject towards a particular configuration.

5.4.3 Results and Discussion

Data was collected from the same 9 subjects (23 - 39 years old) that participated in the previous experiment, but with subjects labeled as C01 through C09 in this experiment. None of the subjects requested any additional rest time between trials. Each subject performed a set of 24 trials, 12 under bang-bang and 12 under proportional signalling condition. The two different signalling conditions were randomly distributed within the sequence of trials and subjects were not informed which signalling condition was in use during each trial. The four different target window configurations listed in table 5.5 were randomly and equally distributed within the sequence of 24 trials per subject. Of the total 216 trials recorded across all subjects, only 9 trials were repeated attempts. As with the previous experiment, the average frame rate for the entire control loop was 22 fps.

5.4.3.1 Qualitative Observations

A number of qualitative observations regarding the subjects' response to the interface and task performance made during the experimental sessions are important to note. These observations aid in the interpretations of the graphical results provided in the following sections. Amongst the nine subjects: seven used a straight-arm posture

during the trials, while the remaining two used a slightly bent posture at the commencement of the trials. Reaching posture variation was entirely up to the choice of the subject. This trend towards subjects selecting the simplest configuration of arm posture is expected. In the previous experiment those subjects that used a moderately bent or wing posture tended to have a higher task completion time on those trials. The experiment operator was able to observe confounding of allocentric and egocentric frames of reference when the subject tried to resolve motion cues if the medial line axis of the forearm was not approximately perpendicular to the frontal plane of the subject.

As with the previous experiment, the reaching task distance of 60cm required the subjects to take at least one step forward to complete the task, with the norm being two or more steps. Several subjects exhibited some moderate timidity in moving forward during the depth tracking phase(s) of the reaching task, but overall subjects seemed to have establish a reasonable level of comfort with unsighted forward movement, which was greater than comfort level observed in the previous experiment. Within the trajectory plots captured during this set of trials, there was substantially less evidence of bounce deviations, caused by jarring steps forward, than in the previous experiment. This seems to indicate the stable kinesthetic position control of the end-effector is a readily learned motor skill developed from proper coordination with lower body movement.

However, the false-backward behaviour manifest of a strong ulnar deviation in wrist posture persisted within this set of trials. Of the 216 trials, 41.7% exhibited this type of false-backward behaviour in the trajectory plots. This is in comparison to the the 39.8% of trials exhibiting the same trajectory plot behaviour in the previous experiment.

These qualitative observations of wrist posture deviations indicate that the wrist posture instability continues to be a compelling issue in the control of an eye-in-hand configuration, even after additional practice with the system through a successive set of trials, or variation of the DES-Controller parameters. Given the available data, it is not feasible to determine why subjects seem to be able to increase their proficiency in stabilized horizontal and vertical position control during forward motion, but there is no increase in proficiency of postural control at the wrist under the same number of cumulative trials.

5.4.3.2 Trajectory and Performance Analysis
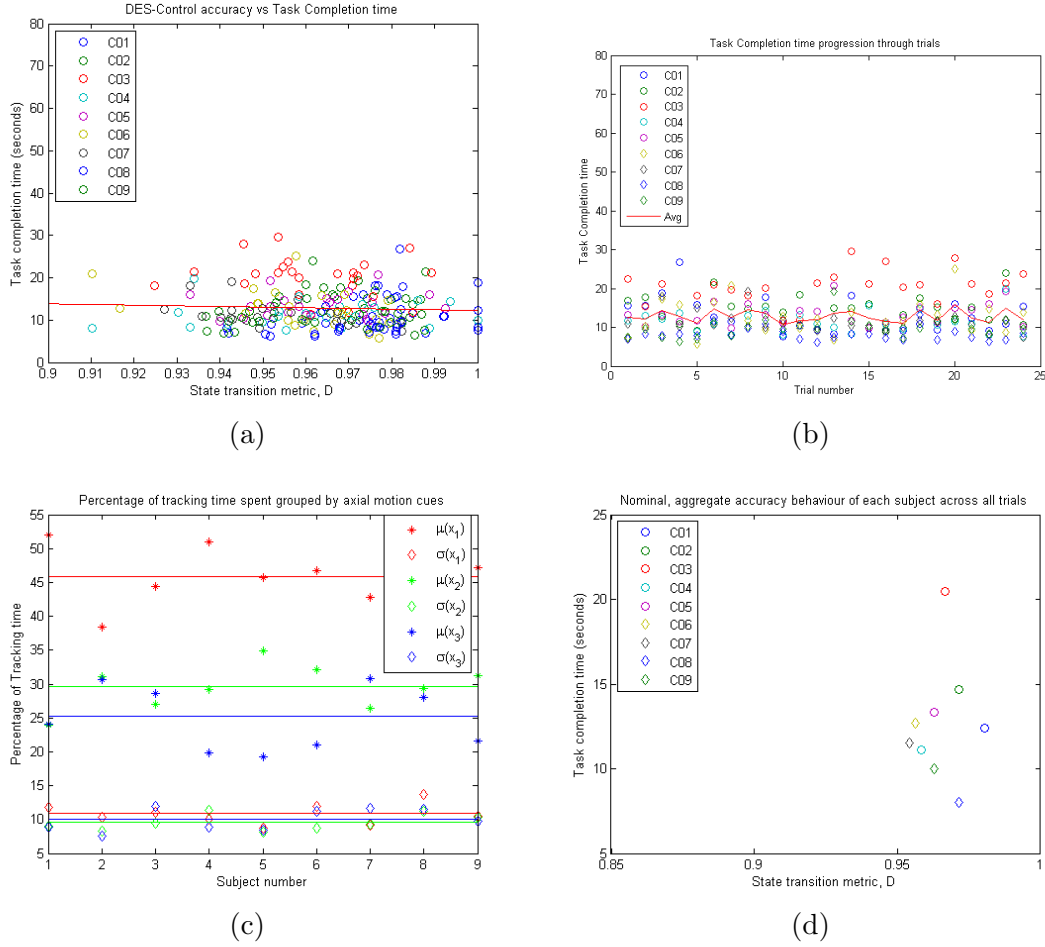
Figure 5.14: DES-Controller state transition accuracy metric and completion time progression across the sequence of trials: (a) Task completion time versus state transition metric for each trial; (b) Task completion time versus trial progression for all subjects; (c) Tracking time statistics by axial movement; (d) nominal aggregate DES metric for each subject

Figure 5.14(a) provides a scatter plot of the recorded task completion time versus the calculated state transition accuracy metric, D. The figure shows a trend of an increase in nominal, aggregate accuracy as expected, with a greater mean accuracy ($\mu(D) = 0.9644$) and lesser standard deviation ($\sigma(D) = 0.0172$), due to the implementation of the hysteresis region. The increase cannot necessarily be fully attributed to the hysteresis region between the inner and outer hypersurfaces. It is reasonable that some portion of the performance increase is due to learning or practice effects.

It is important to note that the data point cluster is much more tightly packed that the similar plot in the previous experiment with a range of $0.91 \leq D \leq 1$ and all reaching task trials completed in under 30 seconds. This is in comparison to an accuracy metric range of $0.82 \leq D \leq 0.98$ with 8.8% of task times exceeding 30 seconds. (see figure 5.11(a))

Figure 5.14(b) plots the task completion times for all trials by trial progression. In contrast to the previous experiment, the averaged subject performance remains relatively constant throughout the progression. This would tend to indicate the most of the subjects have already attained their maximum bound on tracking speed. This should be considered a positive result because it indicates that users readily become comfortable with using the system efficiently.

The plot in figure 5.14(c) shows the mean and standard deviation of percentage tracking-time per movement axis across all trials. The Root-Mean-Squared (RMS) values for mean tracking-time (dashed-line) and standard deviation of tracking-time (solid-line) are included in the plot. The large difference in RMS of means in conjunction with the very similar RMS of standard deviations indicates a bias in the percent tracking-time across the axial sets of motion cues. This is to be expected given that the hybrid-control scheme is prioritized to correct tracking error in x_1 (horizontal), then x_2 (vertical), then x_3 (depth). These statistics in comparison to the similar set in the previous experiment (see figure 5.11(b)) shows the same relative proportions to total tracking-time and hence a consistent behaviour in the CT-plant (human) response to the DES-Controller even under the differing DES-Controller parameter values.

The plot in figure 5.14(d) shows the nominal state transition accuracy metric for each subject. The aggregate behaviour is calculated as the center of mass of the 24 trial cluster of D scores for each subject. We can see that the various users have very similar behaviour in terms of accuracy, but quite a broad range of capability in terms of movement speed. The plot shows a reasonable dispersion of nominal accuracy scores for randomly selected subjects after undergoing some learning effect from the previous experiment. After more than 50 cumulative trials across two experimental sessions, sufficient practice should have occurred that we would expect to see nominal accuracy scores that exhibit a narrow range as the users become accustom to perceiving and interpreting inputs from the vibrotactile interface. However, a larger spread in the nominal speed per subject is expected as that is more limited by their individual

motor capabilities. The individual subjects exhibited nominal behaviours that vary from high-speed / high-accuracy to slow-speed / medium-accuracy to medium-speed / medium-accuracy.
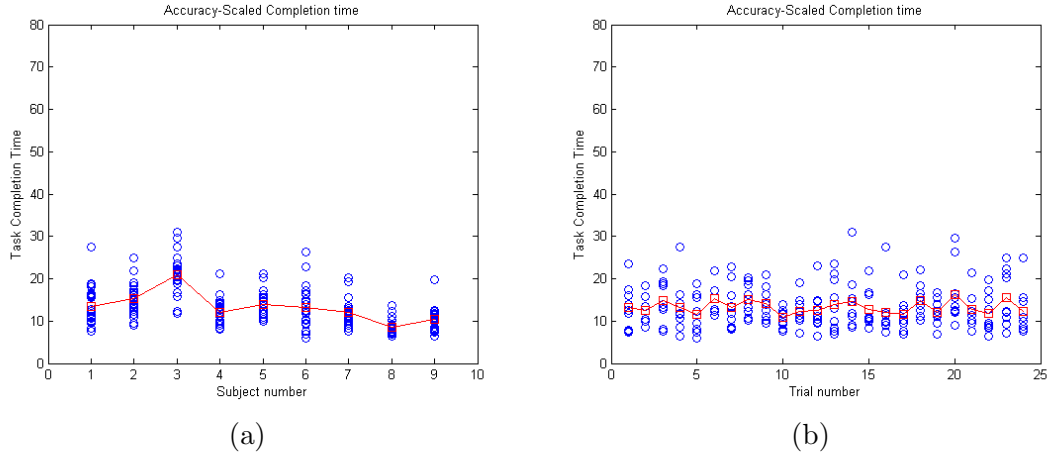
Figure 5.15: Task completion time graphs: (c) scaled task completion time grouped by subject; and (d) scaled task completion time plotted by trial progression.

Given that completion time for a reaching task can be highly variable due to many factors, not the least of which is cumulative learning effect, we present the accuracy-scaled completion time in figure 5.15(a) ordered by subject, and figure 5.15(b) ordered by trial number. The accuracy-scaled completion time for each trial is calculated by dividing the measured completion time by it's corresponding state transition accuracy score, D. The ordering by subject in figure 5.15(a) shows that while substantial performance differences can occur between different plants (users) while preforming the same set of guided reaching tasks, a comparison to the similar plot from the previous experiment (see figure 5.11(c)) demonstrates a much tighter bound on completion time values for almost all subjects. Only subjects C08 and C09 seem to have started at, and maintained their individual maximal task completion speeds across both experimental sessions. Thus those two subjects in particular demonstrate a higher motor acuity than the average behaviour of the entire group.

Upon examining figure 5.15(b) in comparison to its similar plot from the previous experiment (see figure 5.11(d)) we see a much tighter bounds for accuracy-scaled completion time through the progression of trials, as well. This indicates that nomi-

nal performance in both speed and accuracy is converging to the each user's maximal ability limiting the expectation of further performance gains due to a learning effect in subsequent trials of the same task.

As with the previous experiment, we generate the average precision metric across all the trajectory segments within a task trial. To do so we plot the mean DEF ($-1 \leq \theta_e \leq 1$) for each trial. Figures 5.16(a) and 5.16(b) give the mean DEF plotted by subject and by trial progression, respectively. Each point within the plot gives the averaged degree of precision of end-effector motion driven by the CT-plant (internal joint-space controller). When comparing to the corresponding plots from the pervious experiment we can see a general trend for a moderate increase in average precision for the subjects, combined with a moderate decrease in variability of mean DEF scores within each subject. This is an expected result due to cumulative learning effect after many trials. With successive practice in attempting smooth, consistent movement in response to motion cues, subject should tend to maximize the mean DEF per trial. With the data available, it is unclear if and at what range of values a bounded maximum mean DEF would exist for normal human motor response. This is discussed further in the future works presented in Chapter **??**.
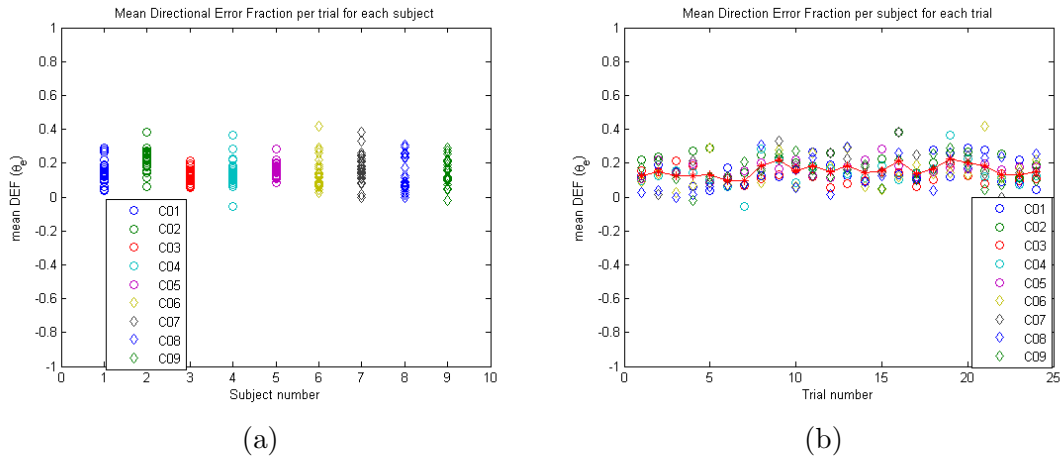
Figure 5.16: Plots of mean directional error fraction (θ_e): (a) mean DEF by subject, and (b) mean DEF by trial progression.

Comparing to similar plots from the previous experiment we see no significant difference in the precision of generalized behaviour of all the subjects. Unlike the appreciable improvement in performance seen in the movement accuracy, the movement

precision is relatively constant. This result is consistent with the results presented within Experiment 2 (section 5.2). In that experiment the results show consistent movement precision with and without the visual reference input signal. So the consistent degree of movement precision between the trials from Experiment 3 and this experiment is reasonable.

(a)

(b)

(c)
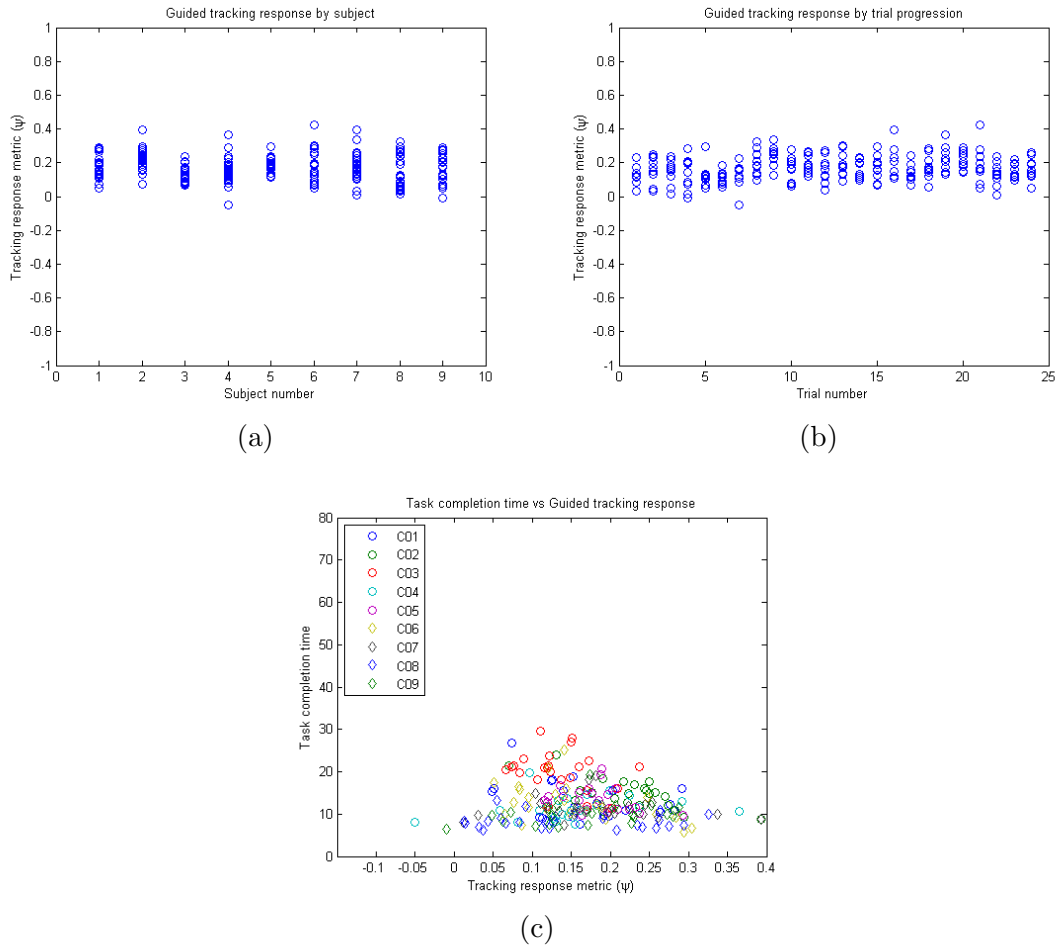
Figure 5.17: Plots of guided reaching task tracking response ψ: (a) by subject, (b) tracking response by trial, and (c) task completion time versus tracking response.

Combining the state transition metric of accuracy and the DEF metric of precision, we can generate a per segment tracking response metric ($-1 \leq \delta[n] \cdot \rho_x[n] \leq 1$) that provides a measure of how well the plant (user) has followed the control (motion cues) per trajectory segment. Then the guided reaching task tracking response is the mean value of segment based metric across the entire trajectory with a range,

$(-1 \leq \psi \leq 1)$. Figures 5.17(a) and 5.17(b) give the guided reaching task tracking response metric by subject and by trial progression, respectively. Figure 5.17(c) gives the mean tracking response vs task completion time for all trials.

(a)   (b)
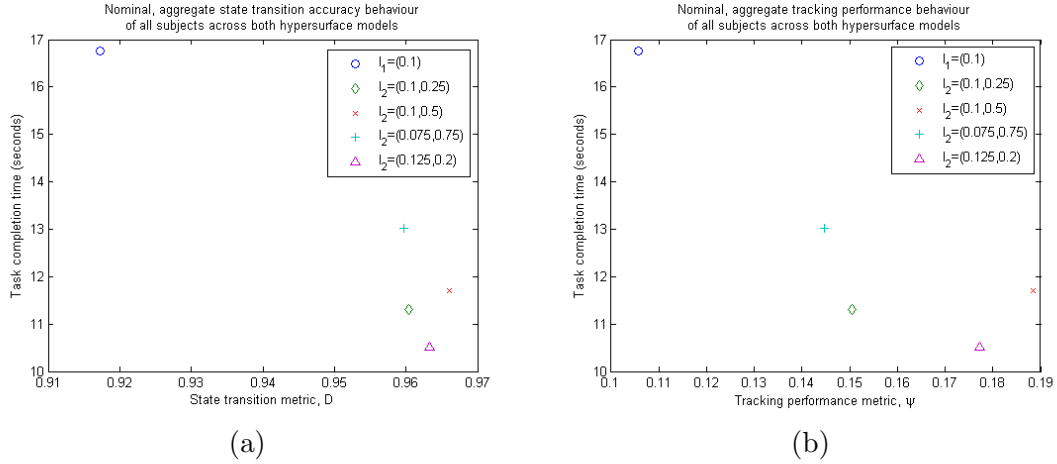
Figure 5.18: The nominal, aggregate metrics for trials under five different DES-model parameter sets: (a) task completion time versus state transition metric;(b) task completion time versus tracking response metric.

To assess the overall effect of DES-Controller parameter variations on the overall system behaviour we plot the task completion time against both nominal, aggregate accuracy metric and the nominal, aggregate tracking task performance metric for the five (l_1, l_2) parameter pairs used in this and the previous experiment. The aggregation of either performance metric describing the nominal behaviour of the generalized user is calculated by finding the center of mass of each cluster of points grouped by model parameter variations shown in figures 5.11(a) and 5.10(a), as well as figures 5.13(e) and 5.17(b).

Figure 5.18(a) provides task completion time versus nominal, aggregate D. The plot shows a significant difference in nominal completion time and state transition accuracy between the l_1-model parameter set and the parameter sets implementing the l_2 model. The l_1-model has a hard boundary for the approximate target region which allows for accidental, incorrect state transitions due to perturbation and minor position drift. In contrast, the points representing the l_2-model parameter sets show discernably greater accuracy and quicker completion time. This result is expected as the l_2-model's hysteresis region was designed to act as a soft boundary, compensating

for the errant state changes occurring from oscillatory crossings due to drift and correction (`Off-Target` to `On-Target`) across one of the single inner hypersurfaces employed in the l_1-model. The constellation of points plotting nominal, aggregate behaviour of the different sets of l_2-model parameters lie in the predicted pattern: from the narrowest approximate target region (or greatest targeting difficulty), to the widest approximate target region (easiest targeting difficulty). This is the expected result.

Figure 5.18(b) provides task completion time versus nominal, aggregate ψ. The plot show that the various sets of DES-Controller parameters used in the current and previous experimental trials maintain their relative position even when examining the tracking performance. This provides confirmation of the consistent development of movement precision demonstrated by the group of subjects over the same set of multiple trials under the various controller configurations. From the plot we see that the constellation of points representing the nominal, aggregate tracking response under the various controller configurations used in this and the previous experiment lie in the predicted pattern, with the narrowest approximate target region (or greatest targeting difficulty) showing the worst performance and an increase in performance through the widening (less difficult) regions.

5.5 Experiment 5: Using Feature Scale to Estimate Depth

This section describes an experiment which demonstrates the effectiveness of using the SIFT scale-space parameter as a camera-to-target depth estimator in a monocular vision based servoing application. The calibration of cartesian distance to the target object in task-space is determined by its relationship in size to a standardized isometric reference object (sphere) of various sizes. The reference object chosen is of a reasonable size for objects that would be grasped by the average user.

Depth estimation is one of the central problems addressed in 3-D reconstruction from multiple images. There is, consequently, substantial literature including texts such as Hartley and Zisserman [100]; Ma *et al.* [129]; and Faugeras and Long [130] that review the subject. Multiple view geometry discussed in these texts shows that several mature solutions exist for the 3-D reconstruction problem. However, when the camera is moving forward along the optical axis, as opposed to a pan or sweep, there are additional complications [131]. Since the planned reach trajectory only requires depth estimation while in a forward motion, these type of solutions do not provide sufficient benefit for the computational complexity and cost. Below we discuss the applicability of some of the other available techniques for depth estimation so that our proposed technique is described within context of the requirements of the guided reaching task.

Depth estimation techniques can be broken down in terms of complexity and accuracy of the system. Some sensing elements such as laser ranger finders and ultrasonic devices can be quite accurate. However, some of the system integration complexities and costs can make them undesirable when a highly accurate estimate is not necessary. As an example, one of the complex difficulties presented from integrating a laser range finder to provide depth is the alignment of the beam. Rigidly mounting the laser range sensor to a fixed point on the hand is not feasible. The mounting point on the glove can easily shift its location on the dorsal side through natural movement of the hand during the reaching task. In addition, the result from Experiments 3 and 4 showed a substantial problem with directional alignment due to the tendency toward ulnar deviation at the wrist. Thus alignment of a depth sensor reference frame at the mount point to any egocentric reference frame on the hand is not assured. This allows for a reasonable likelihood that the ranging data sampled

would be reflected from a surface other than the target, and at a significantly different depth position in the task-space.

A purely image-based solution can come from a stereo vision method using a dual-camera calibrated rig and calculations of correspondence. Complex scene and task-space geometries can be derived from using stereoscopy and Structure from Motion (SfM). That is the technique used in a system such as that proposed by Leibe *et al.* [132]. Their work demonstrates automatic scene geometry estimation from the motion of a stereo camera rig mounted on a vehicle. While this type of technique can produce accurate depth estimation of the target, there are some disadvantages in the context of assistive devices. The main disadvantages are the need for relatively complex multiple view geometry algorithms for 3-D reconstruction, and the need for two cameras instead of one. Monocular vision, in contrast, uses a single camera. However, it needs even more complex multiple view geometry algorithms [133].

As described early on in this work, the accuracy of the depth estimation is not crucial in this class of application. So alternate criteria gains significance. The first being weight. A stereo rig will likely be at least twice as heavy as a monocular vision solution. A stereo rig greatly increases the hardware complexity of the system. Not only is the initial material cost greater, but maintenance is a further issue. A reasonable amount of bumping and/or collisions between the glove and external objects can easily be expected from natural, daily human activity. If the rig loses it's calibrated alignment between the cameras, accurate depth estimation is no longer achievable.

A work advocating the use of scale-space feature detection (SIFT implementation) for eye-in-hand visual servoing applications was reported by Hoffman *et al.* in [134]. Their work reported results of experiments testing the ability of a 5-DOF KATANA manipulator to visually servo to a goal pose using visual control features calculated from the moments of SIFT features (keypoint location, scale, and keypoint orientation). One of the visual control features they propose is $f_{z\sigma} = \frac{1}{n}\sum_{i=1}^{n}\sigma_i$, which drives the translational motion, T_z, along the optical axis. They also state (pg.4265) that the actual distance, z, can be recovered from the scale parameters, σ_i, under the assumption that the initial distance, $\hat{z}$, at the reference frame is known. This is accomplished using the average of the relative scale ratios: $z = \frac{\hat{z}}{n}\sum_{i=1}^{n}\frac{\sigma_i}{\hat{\sigma}_i}$. Basing the distance calculation on the average of the scales for the extracted features poses a significant problem. As the camera moves in T_z additional SIFT features can be incorporated in the extracted set as high frequency image components suddenly

become visible. These emerging features will be of smaller scale and tend to drive the value of the average scale down.

To counter this we propose an alteration to scheme by Hoffman etal by defining the a primary feature, F_p, within the set of extracted SIFT features, where $\sigma_p = \max(\sigma_i)$. The remainder of the features, are referred to as the secondary features. The primary feature represent the physical object as a blob entity in the image; at the largest scale for which the complete object is still discernable after significant Gaussian blurring. The keypoint of F_p will be located at (u_p, v_p); approximately the centroid of the object within the image. Then we can redefine $f_{z\sigma}$ as the weighted average of scales.

$$f_{z\sigma} = a\sigma_p + b \sum_{\substack{i \in \{1..n-1\}, \\ i \neq p}} \sigma_i \tag{5.1}$$

If $a = \frac{1}{n}$ and $b = \frac{n-1}{n}$ then the definition of $f_{z\sigma}$ in [134] is reproduced. Since the primary feature has the greatest significance in determining the size of the object and by extension the distance from the camera, we set $a = 1$ and $b = 0$ to obtain $f_{z\sigma_p}$ sufficient for depth estimation in the guided reaching task. While all extracted features are necessary for the object recognition, not many are representative of the size of physical structures of object. Numerous features can represent textural elements of the object surface. Thus having the secondary features contribute to the average scale can introduce a significant error into the distance estimate.

From equations (3.16) and (5.1) we can define an estimate of the diameter for a boundary circumscribing the projection of the object in the image as a function of primary feature's scale such that

$$d(f_{z\sigma_p}) = 2\sigma_0 f_{z\sigma_p} \tag{5.2}$$

The constant values of 2 is incorporated because $f_{z\sigma_p}$ is representative of the radius, and not the diameter, of the object blob entity; and σ_0 is necessary because Lowe performs an initial Gaussian smoothing of $\sigma_0 = 1.6$ before constructing the scale-space in the implementation of his SIFT algorithm [91].
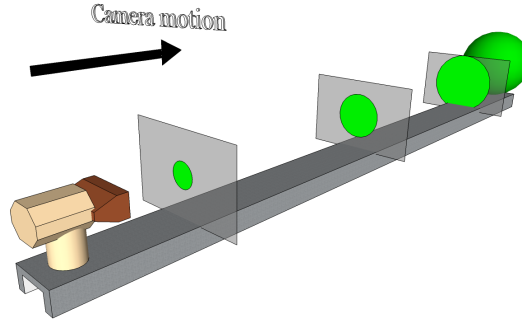
Figure 5.19: Illustration of an isometric object projected onto the image plane

Table 5.6: Intrinsic camera parameters

Focal Length: f	$[948.96830, 951.49261]$
Principal point: $[p_x, p_y]$	$[308.78603, 197.34592]$
Skew: α_c	0
Distortion: k_c	[0.42513, 0.24572, -0.02011, 0.00549, 0.0000]

5.5.1 Experimental Apparatus

Prior to capturing the reference images of the objects, a camera calibration was performed using an 11-inch by 11-inch planar checkerboard. The procedure for performing the camera calibration, and the Matlab toolbox used can be found at [135]. The intrinsic camera parameters extracted from the calibration process: focal length (f_c), principal point (P), skew (α_c), and distortion coefficients (k_c); are provided in Table 5.6.

Figure 5.19 provides and illustration of the experimental rig that was used to simulate the depth tracking and capture of the images for this experiment.

5.5.2 Experimental Procedure

A series of images of a reference object were captured to extract the $\max(\sigma)$ from the keypoint located at the centroid of the object in the image with a feature descriptor representing the entire object as a blob feature. The reference objects selected were items that have a standardized sizes and an isometric shape: a squash ball (40mm

diameter), and a Racquetball (57mm diameter). When the spherical objects are viewed with their significant markings occluded from the camera, they are immune to distortion due to perspective projection. The uniform texture also provides the benefit of minimizing the number of feature descriptors extracted at the smaller scales. The following procedure was used for each object imaged.

For each set of images the camera was placed at a distance of the 103.5 cm from the object on slotted rail. At each depth iteration an image was captured and a set of SIFT feature descriptors are extracted. Then the camera was moved incrementally closer to object along the rail. The incremental camera movements were consistent translations of 25.4mm (1-inch) as the slots were at precisely machined locations on the rail, allowing for accurate recording of distance to target measurements.

5.5.3   Results and Conclusions

The experimental results presented below were obtained using Matlab 7.1 with Image Acquisition Toolbox version 1.9. The camera used is a Logitech QuickCam Messenger, and images were captured at a resolution of 640x480.
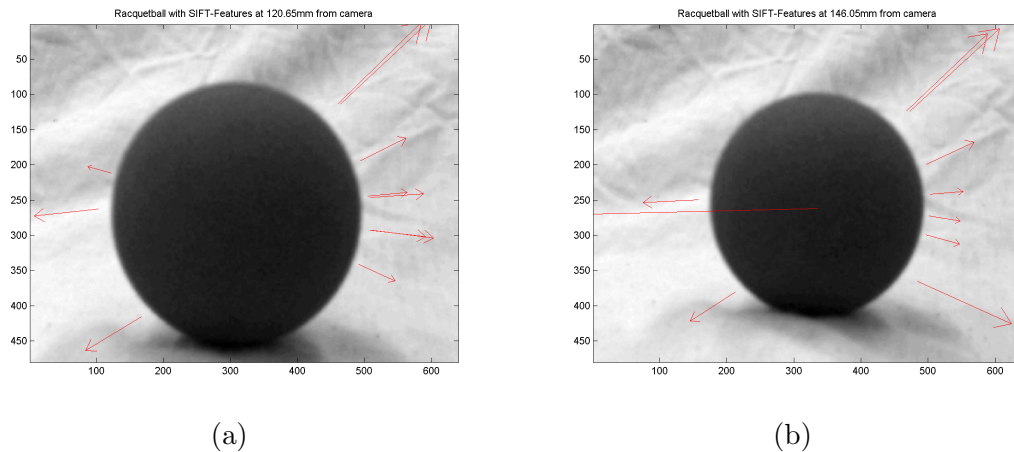
Figure 5.20: Example images of with the scales plotted at the extracted keypoints: (a) racquetball (logo occluded) at $Z = 121$ mm without a primary keypoint; (b) racquetball (logo occluded) at $Z = 146$ mm with a primary keypoint.

Example images are given in Figures 5.20–5.22 illustrating the scale parameter magnitude for SIFT features extracted from the image sets of racquetball and squash-ball objects. The image in Figure 5.20((a)) shows the point in the depth motion of

(a)
(b)
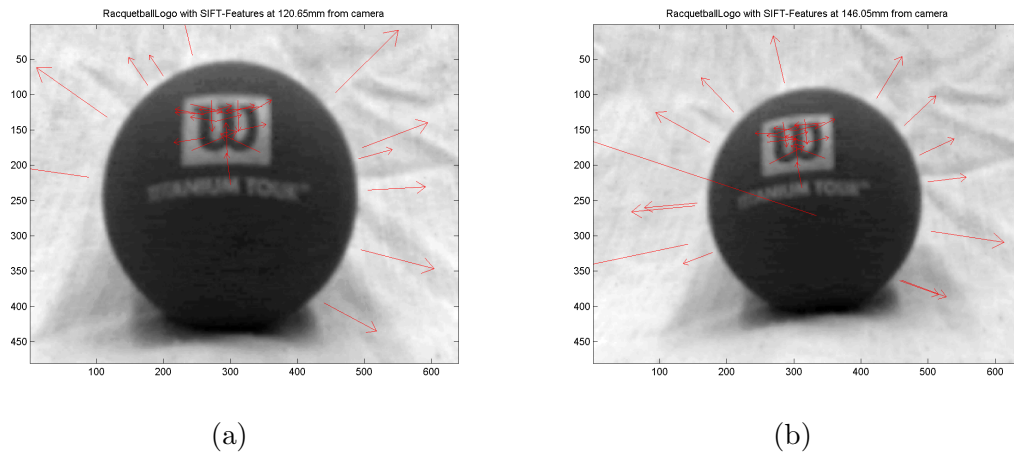
Figure 5.21: Example images of with the scales plotted at the extracted keypoints: (a) racquetball (logo visible) at $Z = 121$ mm without a primary keypoint; (b) racquetball (logo visible) at $Z = 146$ mm with a primary keypoint.

(a)
(b)
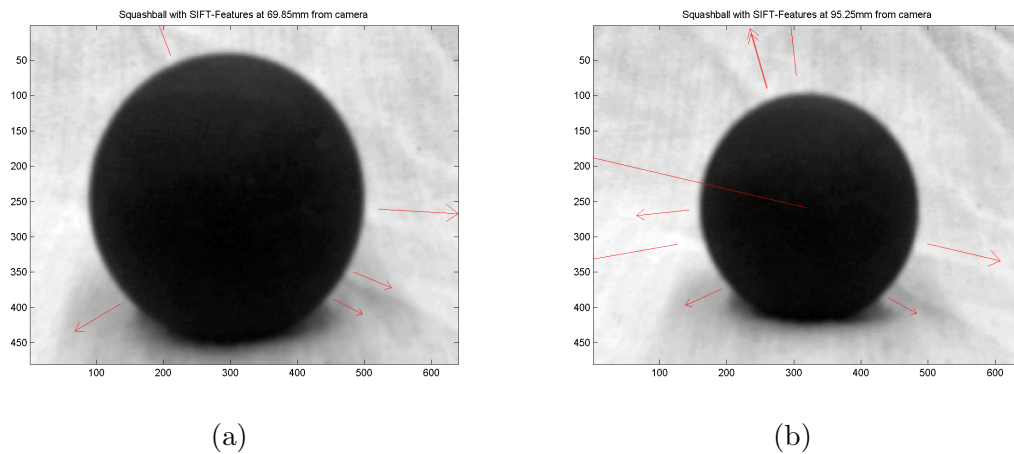
Figure 5.22: Example images of with the scales plotted at the extracted keypoints: (a) squashball at $Z = 69.85$ mm without a primary keypoint; (b) squashball at $Z = 95.25$ mm with a primary keypoint.

images where the primary keypoint is lost due the proximity of the object boundary to image frame boundary. The primary keypoint does not survive the extrema detection and keypoint localization in SIFT after the down sampling is performed to move up an octave in the scale-space. Figure 5.20((b)) shows the SIFT features extracted for the image frame where the primary keypoint was last detected. Figures 5.21((a)) and ((b)) show a similar example for the racquetball with the logo visible.

The primary keypoint for the two orientations of the raquetball object are clearly visible in Figures 5.20((b)) and 5.21((b)). The images also shows the keypoint locations of the secondary features of smaller scale detected at the edges of the raquetball. It is the averaged contribution of these secondary feature keypoint scales that will erroneously alter the distance estimate from the camera to object.

(a)

(b)

(c)
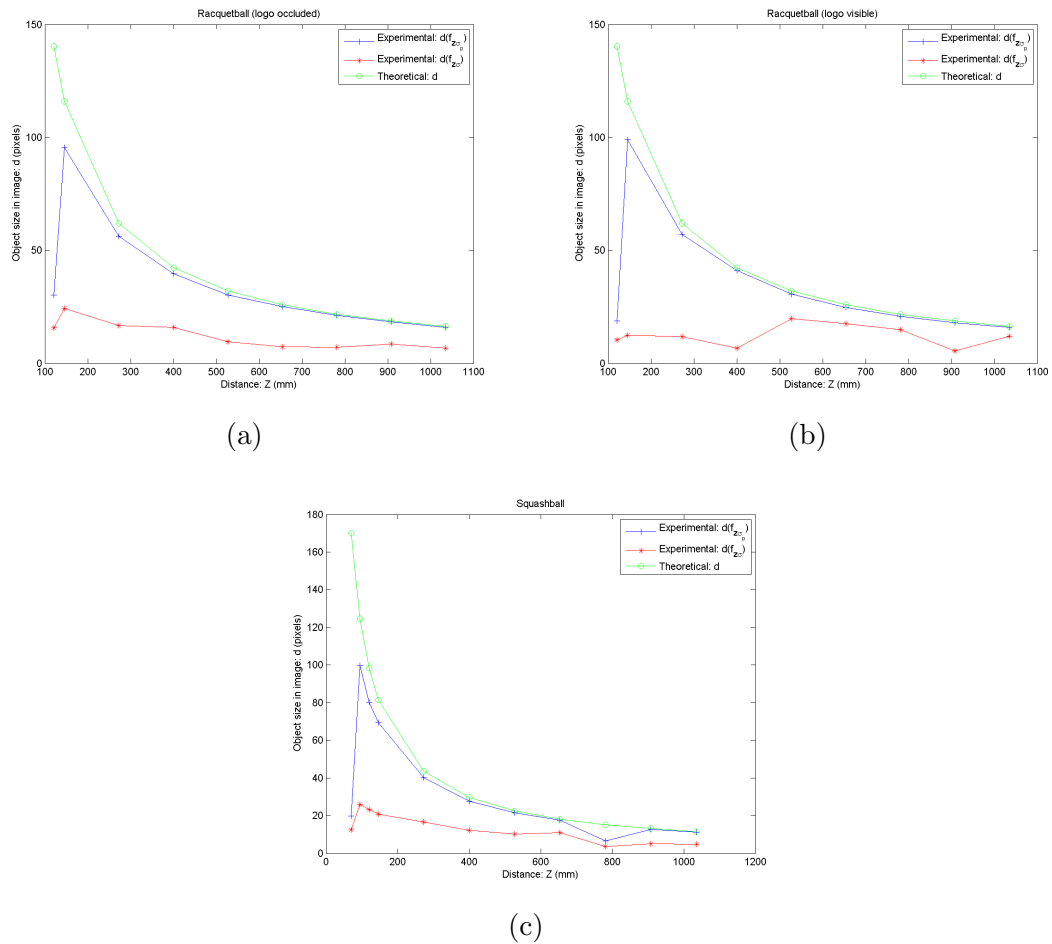
Figure 5.23: Comparison of the accuracy for distance estimation using feature scale between the a conventional moment method and our proposed method: (a) racquetball (logo occluded); (b) racquetball (logo visible); (c) squashball.

Plots comparing the accuracy of the our proposed method described by equation (5.1) and the method proposed in [134] are given in Figure 5.23 relating the distance to target to the magnitude of scale parameter moment. Figure 5.23(a) provides the magnitude estimate for the series of images captured with the logo occluded, and

Figure 5.23(b) provides the magnitude estimates for the series with the logo visible. In each graph the green plot, d, provides the ideal estimate of the object diameter within the image versus the known Z distances from the target object to the calibrated camera (see equation (3.16)). The graphs also provided the estimated pixel diameter of the object image using the method proposed (red plot) by Hoffmann *et al.* [134] and our proposed method (blue plot) with $a = 1$ and $b = 0$. Both graphs show significantly better agreement between our method and the ideal. A similar level of agreement is shown for the squashball object as well. Our method does show a large deviation when the camera is sufficiently close to the object that the image frame boundary interferes with the object boundary, preventing the target object from being recognized as a single blob entity. That is point in depth motion that the original primary feature is lost. As a consequence the estimate of object size is calculated based on the next largest scale feature available, which is only representative of the physical size of that particular feature.

A notable result with our method of heavy weighting towards the primary feature is that the size estimate has a much smoother monotonic behaviour. This is particularly evident in Figure 5.23(b) as it plots the result for target with the logo visible, which produces a substantial number of higher frequency features as the depth motion progresses towards the object. The monotonically increasing function is extremely beneficial for a human plant, as opposed to a conventional robot, since it prevents confusion during perception of the vibrotactile motion cues. If a user is intermittently given feedback that indicates the target object is further away, even though they are aware that they are moving forward, it could cause significant confusion and frustration which could degrade the overall performance of the reaching task.

With this results it can be seen that, given a reference SIFT feature set for a particular target object, a subset of those features can be used to estimate a sufficiently accurate camera to target distance for the generation of forward motion cues. Unlike in purely robotic applications, where the entire reference SIFT feature set would be use to calculate required end-effector pose, it is impractical and unnecessary in this case. As discussed in Chapter 1, the goal of the control is to achieve a close proximity to the target object and allow the user to probe the new tightly constrained task-space to determine how best to contact and grasp the object.

5.6 Experiment 6: Vibrotactile Interface Speed vs Accuracy Characterization

In the previous tracking experiments, sections 5.3 and 5.4, subjects were instructed to follow motion cue given at a quick but comfortable pace. This was necessary as there were seven types of distinct types of vibrotactor output signals to interpret and act upon. In this experiment we constrain the task to that of a 1D virtual tapping test to examine the speed vs accuracy trade-off behaviour is the presence of substitution of somatosensory input in substitution of natural human vision.

5.6.1 Experimental Apparatus

The apparatus used consists of the aiReach system using the l_1-model described previously, but with some necessary functional alterations:

- actuation of vibrotactors is limited to only the `Move-Left` and `Move-Right` units;

- signalling variation is expanded to include Low-to-High proportional signalling as well;

- and a hand held task termination switch for indicating confirmation of an `On-Target` state is used by the subject.

The camera to target distance was approximately 1.4m which was sufficient to achieve a sweeping ipsilateral, horizontal movement from mid-line to past neutral position of the shoulder. All trials were conducted with ipsilateral movements from the initial position to the target for consistency.

5.6.2 Experimental Procedure

The experiment employs a 3x4x4 factorial design. The factors are `signal` condition, `amplitude`, and `width`. The levels for each of the those are BangBang, Proportional High-to-Low, Proportional Low-to-High {BB, HL, LH}; and the sets of movement amplitude $\{A_1, A_2, A_3, A_4\}$ and target width $\{W_1, W_2, W_3, W_4\}$ in pixels. Initial camera to target plane distance was 1.4m to keep pixel pitch consistent. The

Table 5.7: Index of Difficulty by movement amplitude and target width pairings

ID	$W_1 = 32$	$W_2 = 24$	$W_3 = 16$	$W_4 = 8$
$A_1 = 150$	2.5	2.9	3.4	4.3
$A_2 = 125$	2.3	2.6	3.1	4.1
$A_3 = 100$	2	2.4	2.9	3.4
$A_4 = 75$	1.7	2	2.5	3.4

values of the various movement amplitudes, target widths, and corresponding index of difficulty are given in Table 5.7

Subjects were given 10 minutes of practice to become accustom to the system before recorded trials began. Subject were given rest periods to prevent bias due to over exertion. Rest periods were given at a minimum interval of every 6 trials, or upon request.

Subjects were instructed that upon commencement of the motion cue to move as quickly as possible towards the target and then stop movement once vibrotactor output ceases. At that point they were to press the thumb switch to indicate that the `On-Target` condition was achieved. Subjects were further instructed that if they overshoot the target, the opposing vibrotactor would actuate indicating the back tracking was necessary to achieve `On-Target` and that tracking must continue until the no vibrotactor signal is present. Three trials for each `<signal>` x `<amplitude>` x `<width>` combination were recorded, producing 144 trials per subject.

Subjects were aligned with the appropriate nominal amplitude index point prior to the start of each trial all while the subject's vision of the target field was occluded.

5.6.3 Results and Discussion

Results were collected from 3 subjects designated D01 through D03 (2 male and 1 female), ranging in age from 25 to 31 years of age. Due to the inherent, natural position registration instability discussed in experiment 2, the actual initial amplitude of movement for each trial varies slightly. The actual recorded movement amplitude as opposed to the nominal initial amplitude was used in the analysis below. Each subject performed 144 trials which provides a total of 432 trials for analysis.

The actual initial amplitude of each trial is used in the calculation of Index of Difficulty (ID). The total Movement Time (MT) was taken from the initiation of

vibrotactor output until the actuation of the task termination signal from the thumb switch occurs. The most widely accepted 1D Fitts' law formulation of $MT = a + b \cdot ID$, where $ID = \log_2(\frac{A}{W} + 1)$, was used in the analysis below.

5.6.3.1 Qualitative Observations

The subjects made the following qualitative observations regarding their preferences and/or perception of the quality of signalling variations. All three subjects rated the `LH` frequency variation as *"irritating"* or *"frustrating"*. Subjects D01 and D03 specified a preference for `HL` signalling and subject D01 stated no preference between `HL` and `BB` signalling.

5.6.3.2 Performance Analysis

The data collect was used as a whole and partitioned in several ways in an attempt to examine the speed-accuracy trade-off exhibited in a Fitts' law performance model. The graphical and numerical results are presented below.

Figure 5.24 presents a set of scatter plots showing the movement time (MT) versus the index of difficulty (ID) from the recorded trials. Each plot also includes a line of best fit with 95% confidence bounds for the linear fit. Figure 5.24(a) provides the plot and fit line for all recorded trials. Figures 5.24(b)–(d) provide the plots and fit lines for the trials related to the `BB`, `HL`, and `LH` signalling conditions, respectively. The corresponding linear regression coefficients, coefficient of determination (R^2), and RMSE statistics for each are given below:

	b	a	R^2	RMSE
`All`	0.7888	0.9213	0.327	0.9643
`BB`	0.8237	0.7333	0.2059	1.013
`HL`	0.5172	1.729	0.1262	0.8225
`LH`	1.115	0.0475	0.4499	1.029

While the R^2 statistic for the linear regression fit to each of the four sets does not show a high quality fit, an interpretation that supports the qualitative observations from the subjects can be demonstrated by the linear regression coefficients. The slopes of the fit lines used as a general trend in the data shows the steepest slope for the `LH` data set and shallowest slope for the `HL` data set. This indicates a positive

correspondence with the impressions of subjects in that `LH` signalling presented the greatest difficulty in target acquisition and `HL` presented the greatest ease in target acquisition.

(a) (b)

(c) (d)
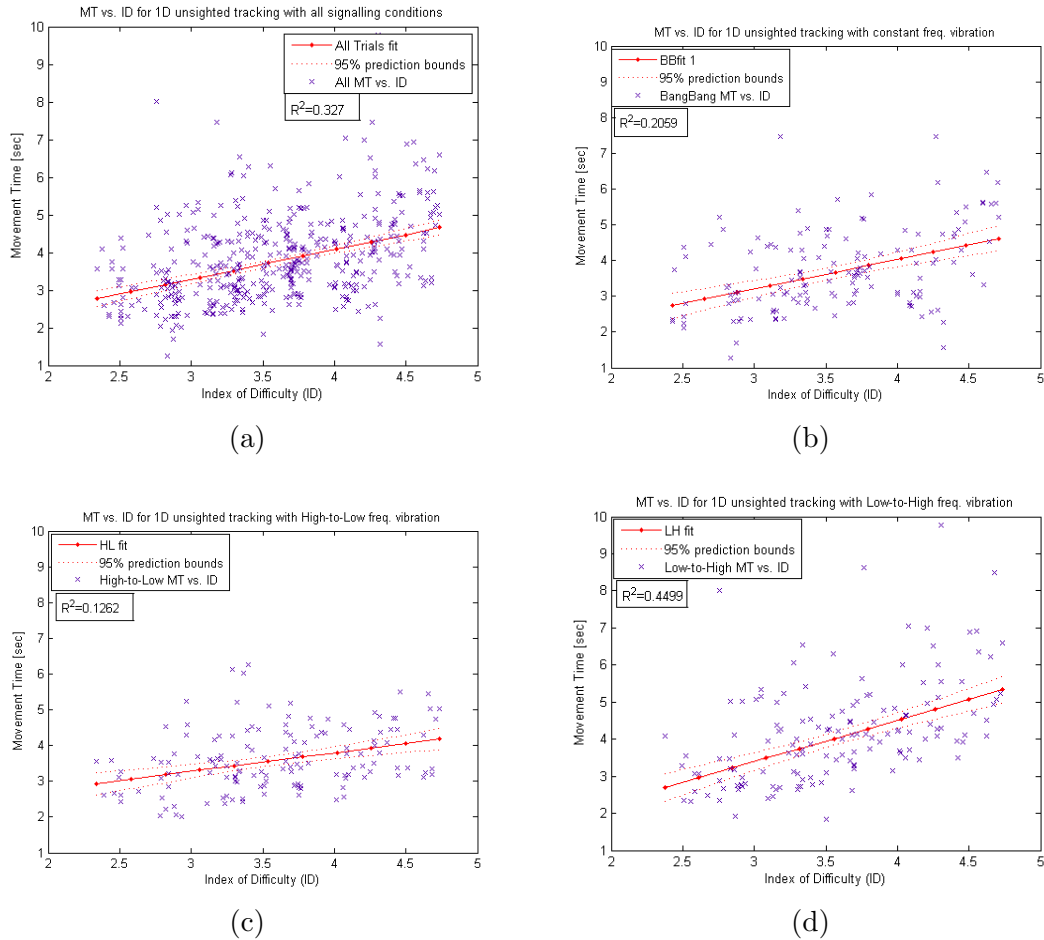
Figure 5.24: Analysis of a 1D Fitts' law performance model of all trials by signalling condition: (a) all trials, (b) all BangBang trials, (c) all High-to-Low trials, and (d) all Low-to-High trials.

Even thought the system makes an attempt at conveying a sense of movement amplitude by proportionally scaling the vibrotactor frequency in comparison, a sense of the target width cannot be readily conveyed to the user. The knowledge of target width is a fundamental component of assessing task difficulty and that lack of a priori knowledge of target width will likely play a strong factor in task performance. The subject does have some perception of target width but through an a posteriori knowledge of the target width if/when they overshoot the far boundary. In that case,

the subject's perception of target width is direct knowledge, but likely imprecise as it will come from kinesthetic perception of their arm position at termination of the ipsilateral vibrotactor signal and actuation of the contralateral vibrotactor. If overshoot does not occur, then there is no direct knowledge of target width. This fundamental difference of input to the task planning mechanism within the human neuromotor control likely creates a significant performance difference from traditional 1D reaching/pointing tasks.

Table 5.8 give the percentage of trials by subject and signalling condition combination that contain one or more overshoot-correction movements before task completion. An examination of the table shows that almost all cases have a considerably high percentage of overshoot-correction movements (above 70%) which indicates that the subjects were having significant difficulty in stopping within initial entry into the target region.

An overshoot-correction movement is defined as passing through the target across the far boundary, then reversing direction in accordance with the contralateral motion cue and returning towards the target. Multiple overshoot-correction movements can occur if the camera passes over the target multiple times before the task termination signal is issued by the subject.

Table 5.8: Percentage of trials exhibiting overshoot-correction movements per subject and signalling condition

	BB	**HL**	**LH**
D01	70.83%	54.17%	70.83%
D02	79.17%	89.58%	79.17%
D03	81.25%	81.25%	87.5%

Figure 5.25 provides the scatter plots of the same data sets as in Figure 5.24 but with the points partitioned into subsets of those trials which contain one or more overshoot-correction movements (green) and those which have none (blue). The plot also includes a linear regression line and 95% confidence prediction bounds to only the points with no overshoot-correction movements. Figure 5.25(a) provides the plot and fit line for all trials without overshoot-correction. Figures 5.25(b)–(d) provide the plots and fit lines for the trials with no overshoot-correction related to the BB, HL, and LH signalling conditions, respectively. The corresponding linear regression

coefficients, coefficient of determination (R^2), and RMSE statistics for each are given below:

	b	a	R^2	RMSE
All	1.05	0.3297	0.1978	1.118
BB	0.463	1.934	0.0432	0.9107
HL	1.159	0.0817	0.176	1.06
LH	1.402	-0.6313	0.1607	1.453

The regression statistics show a very low quality of fit to a linear model, likely due to the fact that majority of points without overshoot-correction movements are clustered towards the lower values of ID. These result from trials with the largest target width, W1, or in some cases the pairings of smaller movement amplitudes with medium-sized target widths. This is likely due to the target width being sufficiently large so that the subject has sufficient time to react (stop) to their perception of an absence of motion cue indicating that they are within the target. However, overall fitting a regression line to these subsets of data does not seem to be an adequate choice of model considering typical Fitts' law tasks do not have such a significant rate of overshoot-correction movements in the task trajectories.

Figure 5.26 provides the scatter plots of the full data set partitioned into subsets by subject. The subject data sets are further partitioned by those trials which contain one or more overshoot-correction movements (green) and those which have none (blue). The plots also includes a linear regression line and 95% confidence prediction bounds to only the points with no overshoot-correction movements. Figures 5.26(a)–(c) provide the plots and fit lines for the trials with no overshoot-correction related to all signalling conditions for subjects D01, D02, and D03, respectively. The corresponding linear regression coefficients, coefficient of determination (R^2), and RMSE statistics for each are given below:

	b	a	R^2	RMSE
D01	1.002	1.261	0.2466	1.042
D02	0.7067	0.4926	0.4868	0.4055
D03	0.7185	0.759	0.3549	0.4214

The regression statics show much better agreement with a 1D Fitts' law performance model on a per subject basis than in the previous sets of figures. This is to be
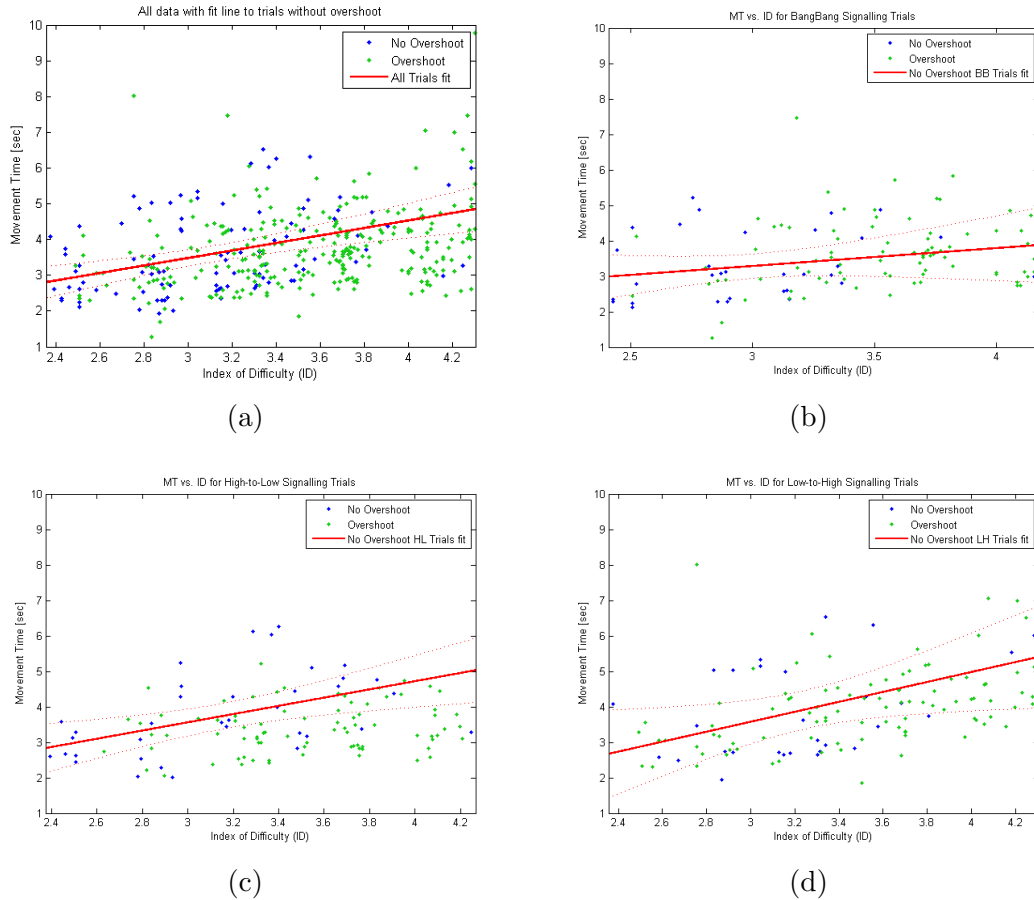
Figure 5.25: Analysis of a 1D Fitts' law performance model of single movement trials by signalling condition: (a) all single movement trials, (b) all BangBang single movement trials, (c) all High-to-Low signal movement trials, and (d) all Low-to-High single movement trials.

expect as the same motor skills capability is applied in each trial. In particular, subjects D02 and D03 show fairly consistent task performance behaviour across the range of task difficulties, demonstrating a greater capability to consistently react to the motion cues at full speed. While the regression statistics do not exhibit the same quality of fit to the linear regression as traditional 1D Fitts' law task experimental results, $R^2 \approx 0.9$ [110, 118, 112], this is also expected. Given the subject is operating in an unsighted condition and has no direct a priori knowledge of the movement amplitude or target width, the feedforward portion of the neuromotor path planning control loop attributed vision is absent. However, the reasonably low dispersion of the data points seems to indicate that some kinesthetic and cognitive (anticipatory reaction to

a change in vibrotactor output) feedforward input does still persist. Examination of the plots shows a strongly linear lower bound for the data points. This would seem to indicate that an expected speed vs accuracy trade-off still persists. The greater variability in the data could be explained by the lesser contribution of feedforward input and the accompanying difference in feedback mechanism. The substitution of somatosensory input in place of vision for the *corrective phase* of the movement is a discrete input in the BB case a very low resolution continuous signal in the HL and LH cases. This is opposition to normal vision being an very high resolution continuous signal representing the relative position of the hand to the target and the size of the target.

(a)

(b)

(c)
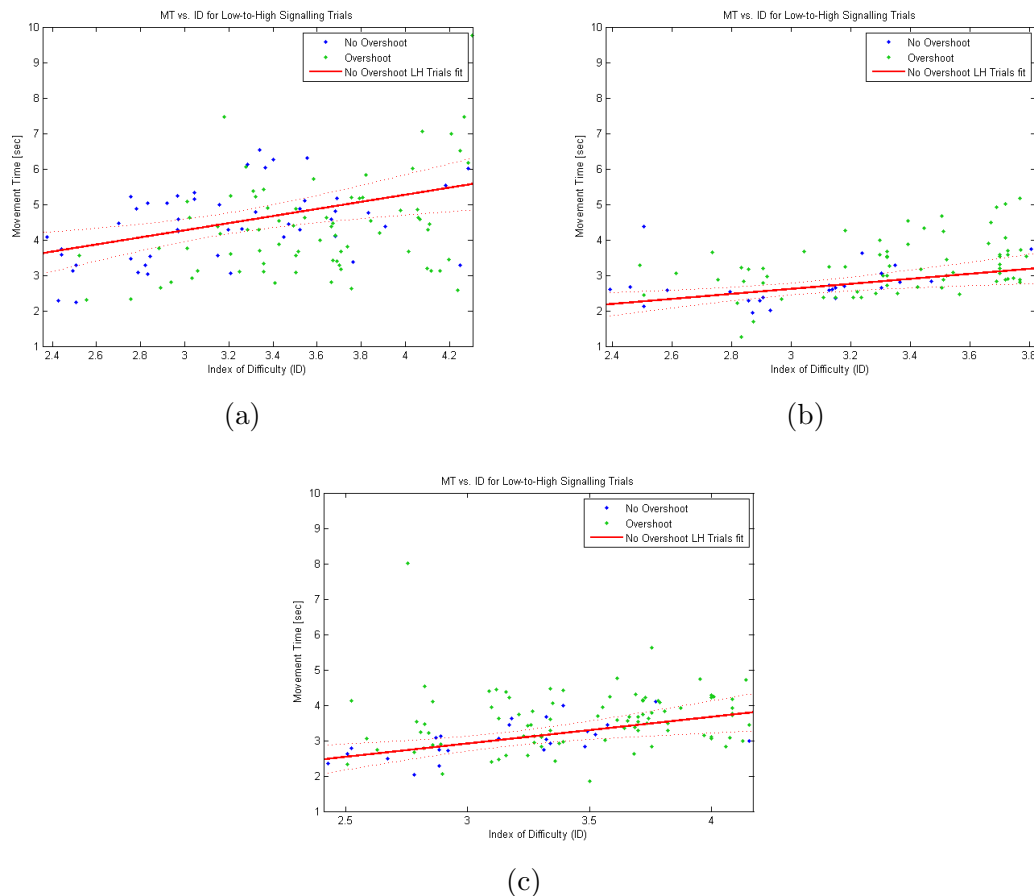
Figure 5.26: Analysis of a 1D Fitts' law performance model of single movement trials by subject: (a) subject D01 trials, (b) subject D02 trials, and (c) subject D03 trials.

Chapter 6 Discussion and Conclusions

In this work a prototype wearable assistive device implementing a hybrid control scheme for object tracking and visual servoing to assist a visually impaired user through a guided reaching task was presented. It was shown that the single camera and vibrotactile interface could provide a viable, indirect *surrogate sight* to stably guide the user's hand toward the target object. However, a significant result of the preliminary investigation was the difficulty in measuring the level of performance of the reaching task. Only the broadest measure of completion with results of either *success* or *failure* could be reliably determined. Even attempting to quantify performance using task completion time was problematic as tracking tasks can very greatly in distance and difficulty. In addition, with the unstructured usability testing of the initial prototype with a generalized subject group, it was found that user intent and conscious action created plant dynamics that made the system seem extremely locally unstable. This was evidenced by user's choosing to perform random movements or ignoring motion cues for a variety of reasons. Yet the system appears to be globally stable as once a deviating plant decided to resume compliance with the motion cues, task completion was inevitable.

6.1   A Novel Model

One of the difficult issues addressed in this work is determining an adequate system model for the guided reaching task movement. The neuromotor movement research community still has competing control models for upper limb movement, which were discussed in the literature review presented in Chapter 2. Thus, even though upper limb movement modeling is a mature area of research the correct system model for a guided reaching is not readily or definitively known. Often the upper limb movement experiments conducted to gather empirical data to support the proposal of either model was done in constrained, two dimensional planar movements scenarios such as along a table top. As such, the results are not entirely applicable to the 3D environment of the free space guided reaching task we investigate within this work.

Instead of pursuing the identification of an adequate continuous-time transfer function for the reaching behaviour, a goal which has been largely illusive to most researchers in the field, we chose an alternate approach that relied on leveraging the intelligence of the user from a task completion perspective. For a guided reaching task the natural high-level model is an IBLM visual servoing model with the user as the joint-space controller. The user's intelligence and tactile sensory capability was leveraged by defining the terminal position as a region adjacent to the target object and not a point on the target object; enabling them to progress to a grasping task using their innate ability to probe and sense the immediate physical environment and the object. Following from that, the controller requirements could be simplified to employ only a single camera in an eye-in-hand configuration and a feature-space controller driven by a three element feature vector of the image space coordinates of the center of mass and a depth approximation feature. This allows for easy interchange of feature extraction and matching techniques of varying complexity.

Since the plant dynamics for the spectrum of users can only be loosely defined, we proposed a feature-space controller that was based on discrete event system. The defined events are consistent across all users (plants) unlike an attempt to achieve control through a conventional kinetic state-space model. Those common events are the achievement of horizontal, vertical, and depth target region alignment of the glove mounted camera. This allows our proposed system to control the process of tracking based on event occurrence and keep the continuous-time plant behaviour separate. The alignment events are defined within image-space so that the system is easily deployed in any reasonable task-space in which the user would venture. In this way, ground truth becomes almost completely unnecessary and the frame of reference is egocentric, which is very natural for the user to interpret.

In this particular implementation, the motion cues that drive the camera motion through the defined tracking events are aligned to the horizontal, vertical, and depth movements, because they are fundamental actions that are discernable through proprioceptive feedback without specific training for the user. However, the hyper-surface boundary locations in the image-space that are the trigger for the tracking events can easily be redefined in alternate configurations if desired.

6.2　Necessity for New Performance Measures

It is very difficult to use conventional control system metrics for performance and stability for a system that incorporates a human in the loop because of the high variability from a nominal dynamic that each individual can introduce in experimental results. Aside from the problematic issue of the neuromotor movement community still maintaining competing proposed control models for upper limb movement, the human factors of perception and intention do not integrate well with conventional control system metrics, simply because conventional plant dynamics do not include "intention" or "choice".

These issues were mostly addressed in the HCI research field by the discovery and extensive application of Fitts' Law. However, it was important to determine if Fitts' law prediction of movement time in relation to index of difficulty was an appropriate metric for this case because of the removal of the direct visual reference signal in targeting process. The literature establishes that movements tasks that can be accurately quantified by a Fitts law performance line have the fundamental dynamic characteristic of a dual subtask composition: a ballistic phase and corrective phase.

Our analysis, which was confirmed with experimental results presented in Experiment 3, showed that a Fitts' law performance line was a very poor fit and thus not applicable, A performance metric which still maintained the underlying aspect of a distinction between accuracy and precision in targeting task movements was proposed to allow for the measure of task performance so that we can determine if alterations to the control parameters or interface truly have a positive of negative effect on the system performance.

6.3　Summary of experimental results and their impact

The proof of concept experiment demonstrated the basic validity of an eye-in-hand, single camera IBLM model for a wearable assistive device. However, some resultant observations were also able to illicite the two fundamental questions that defined the subsequent direction of this work:

- how does one incorporate the significant variability between each individual user's perceptual and motor capability in our system model?

- how does one measure task performance so that we can determine if alterations to the system design, proposed as improvements, actually result in true performance gains?

Experiment 2 showed that the precision of position/movement stability (proprioceptive motor limitation) is the same with or without visual feedback, and that the visual feedback is primarily required for the accuracy of position/movement stability. The results also showed evidence that fatigue had no effect in trials under 60 seconds, so the assumption of time-invariant system for the plant was a reasonable within that duration bound.

Experiment 3 showed that the serial application of Fitts' law and Steering law were not appropriate performance models for this type of task. The perceptual limitation of an unknown terminal point for the reaching task seems to be a fundamental difference in affecting the task dynamics. Without a known terminal point, a continuous ballistic movement phase could not be achieved. The analysis of the initial frontal plane movement data showed that the task is not a ballistic reach to a known target, as verified by comparing performance to that predicted by Fitts' model using classical ballistic reach to formulate the index of difficulty. The pervasive ulnar deviation postural issue which resulted in a depth estimate instability preclude the use of Steering law as a performance model because of a high failure rate (approximately 40% of trials) for forward motion. However the analysis of the trajectory data from the trials showed that the proposed state transition accuracy metric and the directional error fraction precision metric could easily be applied to examine the two types of performance separately and in combination. The performance metrics were also able to indicate trends in performance changes such as learning/practice effects.

Experiment 4 showed that the use of the DES-Controller metrics were able to indicate predicted nominal aggregate performance changes as the DES-Controller parameters were altered. It was shown that the inclusion of the hysteresis region allowed the subjects to complete the guided reaching task with greater accuracy and in less total time. By plotting the nominal, aggregate behaviour of all the subjects to show the performance of a generalized CT-plant, it was demonstrated that guided reaching task tracking response increased as DES-Controller parameters (l_1 and l_2)

were altered from the narrowest approximate target region to the widest in the testing set.

Experiment 5 demonstrated that more rigorous and capable feature extraction and matching scale-space techniques such as SIFT could be employed. Results in the literature had shown that other researchers were able to produce an eye-in-hand visual servoing solution using SIFT with robotic manipulator, but with moderately unstable depth estimation due to scale space traversal.

Experiment 6 investigated the speed vs accuracy phenomena in performance of human motor targeting tasks, a theoretical underpinning of Fitts' law, in presence of the vibrotactile interface employed by the system. The initial results from this pilot study showed that even in the absence of human visual perception of the task difficulty the phenomena still persists as a lower bound in the Movement Time vs Index of Difficulty plots.

Chapter 7 Future Work

We intend the development of this system to be an ongoing process. The major development directions we have identified fall broadly into three categories:performance improvements, control system accuracy and user communication.

7.1 Performance Improvements

A likely beneficial addition to the object recognition portion of the software would be to incorporate windowing (focus of attention) scheme into the vision system. The purpose of the windowing would be to limit feature extraction and matching to only a portion of the image. Once the initial tracking has begun, the trajectory is reasonably predictable and a significant portion of the image can be ignored which would be a moderate performance improvement. The resultant increase in the processed frame rate, would be seen at larger distances from the target object. At closer distances, little to no processing speed benefit would be achieved as the target would encompass most of the frame, so the entire frame would need to processed.

7.2 Control System Accuracy

In addition, it would be beneficial to add functionality could back step through the DES-Controller state history. Introducing state memory would allow the system to more readily reacquire a target object that has been identified but is no longer in the field of view due to unstable user movement.

Currently, one of the most important issues related to accuracy is the wrist posture deviations. It would be beneficial to incorporate some mechanism to eliminate or control the deviations, particularly the lateral ulnar deviation which was the most common and severe occurrence. The most straight forward method to prevent postural deviation of the wrist is to augment the glove so that it is extends through a moderately firm sleeve constraining wrist movement. However, that may not be

beneficial in light of reduced comfort and preventing natural wrist movement that may be necessary during the subsequent grasping task.

The percentage-time tracking results in experiments 3 and 4 alluded to the need to test an alternate configuration of tracking priority. The current tracking priority of x_1 then x_2 then x_3 showed a clear artificially induced bias. One question that arose is whether there exists a natural bias in axial tracking effort. As such a extension to the experiment 3 could be performed in which an alternate state transition function, δ, so that $\max(||x_1||, ||x_2||)$ is given priority for axial alignment. Once horizontal and vertical alignment is achieved, then the system would transition to the depth tracking state. Similar analysis of those trials would then be able to establish if a natural bias in tracking effort exists.

The initial approximate target region dimensions were arbitrarily selected as 20% of the image dimensions. The sets of $\{l_1, l_2\}$ parameter variations expanded the approximate target region out to an inner boundary of 25% and outer boundary of 30% of the image dimensions. An extended set of trials encompassing a greater range of approximate target region sizes to determine if, and at what point task performance begins to degrade significantly.

Lastly, the $\{l_1, l_2\}$ parameters were held constant within each set of trials. The effect within task-space is to create a hyperbolically decreasing approximate target region as depth tracking proceeds towards the target. This allows for loose horizontal and vertical alignment initially, but with an increasing degree of precision as the hand approaches the target object. It would be worthwhile to examine effect of variable $\{l_1, l_2\}$ during depth tracking. This would be employed as the parameters being a function of depth offset, $l_i = f_i(x_3)$. Ideally, this would allow the aiReach system to maintain a minimum approximate target region size in task space, related to the nominal instantaneous drift found in Experiment 2, even beyond the corresponding depth scale.

7.3   User Communication

A practical issue with the construction of the prototype was the placement of vibrotactors. Subjects with smaller hands experienced moderate difficulty with accurate perception of motion cues at two of the transducer sites. Without a properly fitting glove, the motion cue signal from vibrotactor on the palm (`Move-Down`) appears

attenuated to the user. This attenuated vibratory amplitude at the transducer site combined with sympathetic vibration carried in the vibrotactor lead wires strung up and around the distal portion of the foreman lead to some perceptual confusion.

Similar issues at the `Move-Left` (anatomical snuff box) vibrotactor were observed as well. In cases of medium and larger hands fitting tightly within the glove, the placement of that transducer would maintain a stable position closer to the proximal side of the first metacarpal, allowing greater signal perception through bone conduction. With smaller hands the loose fit would allow the glove to shift forward, aligning the transducer contact point with the loose skin between the thumb and index finger; attenuating the perceived signal amplitude. The obvious and simple solution in future iterations is to provide various glove sizes to achieve an properly sized fit for each user. However, this observation has also illuded to further experimentation and analysis of ideal placement site for the vibrotactors. While certain positions may take advantage of the higher perceptual SNR through bone conduction, they may also present detrimental ergonomic affects. Chronic injury due to sustained or long term successive usage may become a factor.

References

[1] M. Jordan and D. Rumelhart, "Forward models - supervised learning with a distal teacher," *Cognitive Science*, vol. 16, pp. 410–416, 1992.

[2] S. Ram and J. Sharf, "The people sensor: a mobility aid for the visually impaired," *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pp. 166 – 7, 1998.

[3] S. Meers and K. Ward, "A vision system for providing the blind with 3D colour perception of the environment," in *Proceedings of the Asia-Pacific Workshop on Visual Information Processing*, (Hong Kong), December 2005.

[4] L. Parker, "Current state of the art in distributed autonomous mobile robotics," in *International Symposium on Distributed Autonomous Robotic Systems*, (Knoxville, TN), pp. 3–12, October 2000.

[5] D. J. Jacques, R. Rodrigo, K. A. McIsaac, and J. Samarabandu, "An object tracking and visual servoing system for the visually impaired," in *Proceedings of the IEEE International Conference on Robotics and Automation*, (Barcelona, Spain), pp. 3510–3515, April 2005.

[6] D. Jacques, R. Rodrigo, K. McIsaac, and J. Samarabandu, "An application framework for measuring the performance of a visual servo control of a reaching task for the visually impaired," in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 894 –901, oct. 2007.

[7] M. Mattar, A. Hanson, and E. Learned-Miller, "Sign classification using local and meta-features," in *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, p. 26, june 2005.

[8] P. Silapachote, J. Weinman, A. Hanson, M. Mattar, and R. Weiss, "Automatic sign detection and recognition in natural scenes," in *Computer Vision*

and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, p. 27, june 2005.

[9] J. Sudol, O. Dialameh, C. Blanchard, and T. Dorcey, "Looktel - a comprehensive platform for computer-aided visual assistance," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 73 –80, june 2010.

[10] X. Chen and A. Yuille, "Detecting and reading text in natural scenes," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II–366 – II–373 Vol.2, june-2 2004.

[11] I. Ulrich and J. Borenstein, "The guidecane, applying mobile robot technologies to assist the visually impaired," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 31, no. 2, pp. 131–136, 2001.

[12] S. Shoval, I. Ulrich, and J. Borenstein, *Computerized Obstacle Avoidance Systems for the Blind and Visually Impaired*. Intelligent Systems and Technologies in Rehabilitation Engineering, CRC Press, 2000.

[13] R. Audette, J. Balthazaar, C. Dunk, and J. Zelek, "A stereo-vision system for the visually impaired," Tech. Rep. 2000-41x-1, School of Engineering,University of Guelph, Guelph, ON, Canada, December 2000.

[14] J. Coughlan and R. Manduchi, "Functional assessment of a camera phone based wayfinding system operated by blind and visually impaired users," *International Journal of Artificial Intelligence Tools, Special Issue on Artificial Intelligence Based Assistive Technologies: Methods and Systems for People with Disabilities*, vol. 18, pp. 379–397, Jun. 1 2009.

[15] H. Hile, R. Grzeszczuk, A. Liu, R. Vedantham, J. Koecka, and G. Borriello, "Landmark-based pedestrian navigation with enhanced spatial reasoning," in *Pervasive Computing* (H. Tokuda, M. Beigl, A. Friday, A. Brush, and Y. Tobe, eds.), vol. 5538 of *Lecture Notes in Computer Science*, pp. 59–76, Springer Berlin / Heidelberg, 2009.

[16] K. Zawrotny, A. Craig, D. Weiser, R. Klatzky, and G. Stetten, "Fingertip vibratory transducer for detecting optical edges using regenerative feedback," in *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2006 14th Symposium on*, pp. 373–374, march 2006.

[17] G. Stetten, R. Klatzky, B. Nichol, J. Galeotti, K. Rockot, K. Zawrotny, D. Weiser, N. Sendgikoski, and S. Horvath, "Fingersight: Fingertip visual haptic sensing and control," in *Haptic Audio Visual Environments and their Applications*, (Ottawa, Canada), pp. 80–83, IEEE International Workshop, 12-14 October 2007.

[18] H. Watanabe, S. Kotani, N. Kiyohiro, and S. Hashiguchi, "Research and development of wearable walking mate system," in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 621 –625, oct. 2007.

[19] D. Yuan and R. Manduchi, "Dynamic environment exploration using a virtual white cane," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 243–249,, june 2005.

[20] Y. Kaneko, T. Harada, Y. Hirahara, Y. Kikuchi, S. Yamada, K. Yanashima, and K. Magatani, "Development of the navigation system for the visually impaired," *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering 2003 (IEEE Cat. No.03EX711)*, pp. 238–239, 2003.

[21] J. Bigham, C. Jayant, A. Miller, B. White, and T. Yeh, "Vizwiz::locateit - enabling blind people to locate objects in their environment," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 65 –72, june 2010.

[22] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh, "Vizwiz: nearly real-time answers to visual questions," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, (New York, NY, USA), pp. 333–342, ACM, 2010.

[23] E. L. Brady, "Analyzing visual questions from visually impaired users," in *The proceedings of the 13th international ACM SIGACCESS conference on Comput-*

ers and accessibility, ASSETS '11, (New York, NY, USA), pp. 309–310, ACM, 2011.

[24] M. Bergamasco, B. Allotta, L. Bosio, L. Ferretti, G. Parrini, G. Prisco, F. Salsedo, and Sartini, "Arm exoskeleton system for teleoperation and virtual environments applications," in *Proceedings of the IEEE International Conference on Robotics and Automation*, (San Diego, CA), pp. 1449–1454, May 1994.

[25] A. Gupta and M. K. O'Malley, "Design of a haptic arm exoskeleton for training and rehabilitation," *Mechatronics, IEEE/ASME Transactions on*, vol. 11, pp. 280 – 289, june 2006.

[26] J. C. Perry and J. Rosen, "Design of a 7 degree-of-freedom upper-limb powered exoskeleton," in *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pp. 805 –810, feb. 2006.

[27] J. G. Linvill and J. Bliss, "A direct translation reading aid for the blind," *Proceedings of the Institute of Electrical and Electronic Engineers*, vol. 54, pp. 40–51, 1966.

[28] H. Z. Tan and A. Pentland, "Tactual displays for wearable computing," in *Proceedings of the International Symposium on Wearable Computers*, (Cambridge, Massachusetts, USA), October 1997.

[29] H. Z. Tan and A. Pentland, "Tactual displays for sensory substitution and wearable computers," in *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, (New York, NY, USA), ACM, 2005.

[30] P. I. Corke, *Visual Contrl of Robots: High-performance Visual Servoing*. Taunton, Somerset, England: Reserch Studies Press Ltd., 1997.

[31] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Transactions Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.

[32] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimentatl Psychology*, vol. 47, pp. 381–391, 1954.

[33] N. Ferrier, "Achieving a fitts law relationship for visual guided reaching," *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 903 – 910, 1998.

[34] A. Murata and H. Iwase, "Proposal of two-dimensional effective target width in fitts' law," *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*, vol. vol.2, pp. 265 – 270, 1999.

[35] A. Murata, "Extending effective target width in fitts' law to a two-dimensional point task," *International Journal of HCI*, vol. 11, no. 2, pp. 137–152, 1999.

[36] N. Yang, D. Jin, M. Zhang, C. Huang, and R. Wang, "An extending fitts' law for human upper limb performance evaluation," *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.01CH37272)*, vol. 2, pp. 1240 – 3, 2001.

[37] T. Grossman and R. Balakrishnan, "Pointing at trivariate targets in 3d environments," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 447 – 454, 2004.

[38] D. A. Rosenbaum, *Human Motor Control.* Elsevier Academic Press, second ed., 2010.

[39] R. Cohen and D. Rosenbaum, "Where objects are grasped reveals how grasps are planned - generation and recall of motor plans," *Experimental Brain Research*, vol. 157, pp. 486–495, 2004.

[40] D. Elliot, W. Helsen, and R. Chua, "A century later: Woodworth's (1899) two-component model of goal-directed aiming," *Psychological Bulletin*, vol. 127, pp. 342–357, 2001.

[41] K. Ericsson, R. Krampe, and C. Tesch-Romer, "The role of deliberate practice in the acquisition of expert performance," *Psychological Review*, vol. 100, pp. 363–406, 1993.

[42] K. Keetch, R. Schmidt, T. Lee, and D. Young, "Especial skills: Their emergence with massive amounts of practice," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, pp. 970–978, 2005.

[43] M. Merzenich, R. Nelson, M. Stryker, M. Cynder, A. Shoppmann, and J. Zook, "Somatosensory cortical map changes following digit amputation in adult monkeys," *Journal of Comparative Neurology*, vol. 224, pp. 591–605, 1984.

[44] L. Snyder, A. Batista, and R. Andersen, "Coding of intention in the posterior parietal cortex," *Nature*, vol. 386, pp. 167–170, 1997.

[45] G. Rizzolatti, L. Fogassi, and V. Gallese, "Parietal cortex: from sight to action," *Current Opinon in Neurbiology*, vol. 7, pp. 562–567, 1997.

[46] A. Batista, C. Buneo, L. Snyder, and R. Andersen, "Reach plans in eye-centered coordinates," *Science*, vol. 285, pp. 257–260, 1999.

[47] M. DESMURGET, C. Epstein, R. Turner, G. Prablanc, C. Alexander, and S. Grafton, "Role of the posterior parietal cortex in updating reaching movements to a visual target," *Nature: Neuroscience*, vol. 2, pp. 563–567, 1999.

[48] V. Stuphorn, K. Hoffmann, and L. Miller, "Correlation of primate superior colliculus and reticular formation discharge with proximal limb muscle activity," *Journal of Neurophysiology*, vol. 81, pp. 1978–1982, 1999.

[49] V. Stuphorn, E. Bauswein, and K. Hoffman, "Neurons in the primate superior colliculus coding of arm movements in gaze-related coordinates," *Journal of Neurophysiology*, vol. 83, pp. 1283–1299, 00.

[50] J. Gordon and C. Ghez, "Accuracy of planar reaching movements: I. independance of direction and extent of variability," *Experimental Brain Research*, vol. 99, pp. 97–111, 1994.

[51] M. Vince and A. Welford, "Time taken to change the speed of a response," *Nature*, vol. 213, pp. 532–533, 1967.

[52] D. Rosenbaum, "Human movement initiation: Specification of arm, direction, and extent," *Journal of Experimental Psychology: General*, vol. 109, pp. 444–474, 1980.

[53] R. Woodworth, "The accuracy of voluntary movement," *Pyschological Review*, vol. 3, pp. 1–119, 1899.

[54] S. Keele and M. Posner, "Processing visual feedback in rapid movement," *Journal of Experimental Psychology*, vol. 77, pp. 155–158, 1968.

[55] L. G. Carlton, "Processing visual feedback information for movement control," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, no. 5, pp. 1019 – 1030, 1981.

[56] H. Zelaznik, B. Hawkins, and L. Kisselburgh, "Rapid visual feedback processing in single-aiming movements," *Journal of Motor Behaviour*, vol. 15, pp. 217–236, 83.

[57] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion in Neurobiology*, vol. 9, pp. 718–727, 1999.

[58] D. Wolpert, Z. Ghahramani, and M. Jordan, "An internal model for sensorimotor integration," *Science*, vol. 269, pp. 1880–1882, 1995.

[59] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Networks*, vol. 11, pp. 1317–1329, 1998.

[60] P. Morasso, "Spatial control of arm movements," *Experimental Brain Research*, vol. 42, pp. 223–227, 1981.

[61] W. Abend, E. Bizzi, and P. Morasso, "Human arm trajectory formation," *Brain*, vol. 105, pp. 331–348, 1982.

[62] J. Soechting and F. Lacquaniti, "Invariant characteristics of a pointing movement in man," *Journal of Neuroscience*, vol. 1, pp. 710–720, Jul. 1981.

[63] S. Antipolis, "Telecommunications keypads and keyboards; tactile identifier," ES 201 381, European Telecommunications Standardisation Institute (ETSI), France.

[64] J. B. F. van Erp, "Guidelines for the use of vibro-tactile displays," in *Proceedings of EuroHaptics*, pp. 18–22, 2002.

[65] L. A. Jones and N. B. Sarter, "Tactile displays: Guidance for their design and application," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, pp. 90–111, Feb. 2008.

[66] R. T. Verrillo, "Investigation of some parameters of the cutaneous threshold for vibration," *Journal of Acoustical Society of America*, vol. 34, no. 11, pp. 1768–1773, 1962.

[67] R. T. Verrillo, "Temporal summation in vibrotactile sensitivity," *Journal of Acoustical Society of America*, vol. 37, no. 5, pp. 843–846, 1962.

[68] R. T. Verrillo, "Vibrotactile thresholds for hairy skin," *Journal of Experimental Psychology*, vol. 72, no. 1, pp. 47–50, 1966.

[69] C. Sherrick and R. Cholewiak, "Cutaneous sensitivity," in *Handbook of Perception and Human Performance* (K. Boff, L. Kauffman, and J. Thomas, eds.), ch. 12, pp. 1–57, New York, NY: John Wiley and Sons, 1986.

[70] G. Gescheider, "Temporal relations in cutaneous stimulation," in *Cutaneous Communications Systems and Devices* (F. Geldard, ed.), Austin, TX: The Pyschonomic Society, 1974.

[71] L. Petrosino and D. Fucci, "Temporal resolution of the aging tactile sensory system," *Perceptual and Motor Skills*, vol. 68, pp. 288–290, 1989.

[72] J. Craig, "Difference threshold for intensity of tactile stimuli," *Perception and Psychophysics*, vol. 11, no. 2, pp. 150–152, 1972.

[73] G. Goff, "Differential discrimmination of frequency of cutaneous mechanical vibration," *Journal of Experimental Pyschology*, vol. 74, no. 2, pp. 294–299, 67.

[74] K. Johnson and J. Philips, "Tactile spatial resolution. i. two point discrimination, gap detection, grating resolution, and letter recognition," *Journal of Neurophysiology*, vol. 46, pp. 1177–1191, Dec. 1981.

[75] R. Cholewiak and A. Collins, "Vibrotactile localization on the arm: Effects of place, space, and age," *Perception and Psychophysics*, vol. 65, pp. 1058–1077, 2003.

[76] C. Sherrick, R. Cholewiak, and A. Collins, "The localization of low- and high-frequency vibrotactile stimuli," *Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 169–178, 1990.

[77] R. Verrillo and G. Gescheider, "Vibrotactile masking: Effects of one- and two-site stimulation," *Perception and Pyschophyics*, vol. 33, no. 4, pp. 379–387, 1983.

[78] G. Wulf, *Attention and motor skills learning*. Human Kinetics, first ed., 2007.

[79] H. Phong Pham and R. Chellali, "Frequency modulation based vibrotactile device for teleoperation," in *Space Mission Challenges for Information Technology, 2009. SMC-IT 2009. Third IEEE International Conference on*, pp. 98 –105, 2009.

[80] A. Sanderson and L. Weiss, "Image-based visual servo control using relational graph error signals," *Proceedings of the International Conference on Cybernetics and Society*, pp. 1074 – 7, 1980.

[81] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Transactions on Robotics and Automation*, vol. 21, no. 6, 2005.

[82] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[83] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, pp. 63–86, October 2004.

[84] T. Linderberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 2, pp. 225–270, 1994.

[85] T. Linderberg, "Principles for automtic scale selection," tech. rep., Department of Numerical Analysis and Computing Science KTH (Royal Institute of Technology), S-100 44 Stockholm, Sweden., 1998.

[86] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, pp. 151–172, June 2000.

[87] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1615–1630, October 2005.

[88] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, pp. 43–72, November 2005.

[89] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3-D objects," in *Proceedings of the Tenth IEEE International Conference on Computer Vision*, vol. 1, (Washington, DC), pp. 800–807, October 2005.

[90] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, (Kerkyra, Corfu, Greece), pp. 1150 – 1157, September 1999.

[91] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 31–110, November 2004.

[92] R. Rodrigo, Z. Chen, and J. Samarabandu, "Feature motion for monocular robot navigation," in *Proceedings of IEEE International Conference on Information and Automation*, (Colombo, Sri Lanka), p. 5, December 2006.

[93] J. Bigun, *Vision with Direction: A Systematical Introduction to Image Processing and Computer Vision.* Springer-Verlag Berlin Heidelberg, 2006.

[94] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 79–116, November 1998.

[95] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, (Los Alamitos, CA), pp. 593–600, IEEE Computer Society Press, June 1994.

[96] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proceedings of the British Machine Vision Conference*, (Cardiff, UK), pp. 656–665, September 2002.

[97] C. Harris and M. J. Stephens, "A combined corner and edge detector," in *In Proceedings of the 4th Alvey Vision Conference*, (Manchester, UK), pp. 147–152, August 1988.

[98] J. Sivic and A. Zisserman, "Efficient visual content retrieval and mining in videos," in *Advances in Multimedia Information Processing - PCM 2004*, vol. 3332, pp. 471–478, Springer Berlin Heidelberg, 2004.

[99] J. Sivic, B. C. Russel, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their location in images," in *Proceedings of the Eleventh IEEE International Conference on Computer Vision*, (Beijing, China), pp. 370–377, October 2005.

[100] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed., 2003.

[101] J. Stiver and P. Antsaklis, "Modeling and analysis of hybrid control systems," in *Decision and Control, 1992., Proceedings of the 31st IEEE Conference on*, vol. 4, pp. 3748–3751, 1992.

[102] J. Stiver and P. Antsaklis, "Extracting discrete event system models from hybrid control systems," in *Intelligent Control, 1993., Proceedings of the 1993 IEEE International Symposium on*, pp. 298–301, Aug 1993.

[103] J. Stiver and P. Antsaklis, "On the controllability of hybrid control systems," in *Decision and Control, 1993., Proceedings of the 32nd IEEE Conference on*, vol. 1, pp. 294–299, Dec 1993.

[104] J. A. Stiver and P. J. Antsaklis, "State space partitioning for hybrid control systems," in *American Control Conference, 1993*, pp. 2303–2304, June 1993.

[105] J. Stiver, P. Antsaklis, and M. Lemmon, "Digital control from a hybrid perspective," in *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, vol. 4, pp. 4241–4246 vol.4, Dec 1994.

[106] J. Stiver, P. Antsaklis, and M. Lemmon, "Hybrid control system design based on natural invariants," in *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, vol. 2, pp. 1455–1460 vol.2, Dec 1995.

[107] X. Koutsoukos, P. Antsaklis, J. Stiver, and M. Lemmon, "Supervisory control of hybrid systems," *Proceedings of the IEEE*, vol. 88, pp. 1026 –1049, jul 2000.

[108] J. Kalaska, "Reaching movements to visual targets: neuronal representations of sensori-motor transformations," *Seminars in Neuroscience*, vol. 3, no. 1, pp. 67 – 80, 1991.

[109] P. Fitts and J. Peterson, "Information capacity of discrete motor responses," *Journal of Experimentatl Psychology*, vol. 67, pp. 103–112, 1964.

[110] I. MacKenzie, "A note on the information theoretic basis for fitts' law," *Journal of Motor Behavior*, vol. 21, pp. 323–330, 1989.

[111] A. Welford, *Fundamental of Skill*. London: Metheun, 1968.

[112] I. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Human Computer Interaction*, vol. 7, pp. 91–139, 1992.

[113] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of fitts law research in hci," *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 751 – 789, 2004. ¡ce:title¿Fitts' law 50 years later: applications and contributions from human-computer interaction¡/ce:title¿.

[114] R. J. Jagacinski and J. M. Flach, *Control Theory for Humans: A Quantative Approach to Modeling Performance*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2003.

[115] G. Langolf, *Human motor performance in precise microscopic work - Development of standard data for microscopic assembly*. PhD thesis, University of Michigan, Ann Arbor, MI, 1973.

[116] G. Langolf, D. Chaffin, and J. Foulke, "An investigation of fitts' law using a wide range of movement amplitudes," *Journal of Motor Behaviour*, vol. 8, pp. 113–128, 1976.

[117] E. Crossman, *The measurement of perceptual load in manual operations*. Phd thesis, University of Birmingham, 1956.

[118] S. MacKenzie and W. Buxton, "Extending fitts' law to two-dimensional tasks," *ACM CHI*, pp. 219–226, 1992.

[119] E. Hoffmann and I. Sheikh, "Effect of varying target height in fitts' movement task," *Ergonomics*, vol. 37, no. 6, pp. 1071–1088, 1994.

[120] C. Ware and R. Balakrishnan, "Reaching for objects in vr displays: Lag and frame rate," *ACM TOCHI*, vol. 1, no. 4, pp. 331–356, 1994.

[121] J. Accot and S. Zhai, "More than dotting the i's — foundations for crossing-based interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, CHI '02, (New York, NY, USA), pp. 73–80, ACM, 2002.

[122] J. Accot and S. Zhai, "Refining fitts' law models for bivariate pointing," in *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York, NY, USA), pp. 193–200, ACM, 2003.

[123] J. Accot and Z. Shumin, "Beyond fitts' law: models for trajectory-based hci tasks," in *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York, NY, USA), pp. 295–302, ACM, 1997.

[124] N. Friedlander, K. Schlueter, and M. Mantei, "Bullseye when fitts' law doesn't fit," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '98, (New York, NY, USA), pp. 257–264, 1998.

[125] J. J. Buchanan, J.-H. Park, and C. H. Shea, "Systematic scaling of target width: dynamics, planning, and feedback," *Neuroscience Letters*, vol. 367, no. 3, pp. 317 – 322, 2004.

[126] J. Buchanan, J.-H. Park, and C. Shea, "Target width scaling in a repetitive aiming task: switching between cyclical and discrete units of action," *Experimental Brain Research*, vol. 175, pp. 710–725, 2006.

[127] Y. Guiard, "On fitts's and hooke's laws: Simple harmonic movement in upper-limb cyclical aiming," *Acta Psychologica*, vol. 82, no. 1-3, pp. 139–159, 1993.

[128] F. Bonnetblanc, O. Martin, and N. Teasdale, "Pointing to a target from an upright standing position: anticipatory postural adjustments are modulated by the size of the target in humans," *Neuroscience Letters*, vol. 358, no. 3, pp. 181 – 184, 2004.

[129] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision.* Springer-Verlag, New York Inc., 2004.

[130] O. Faugeras and Q.-T. Long, *The Geometry of Multiple Images.* The MIT Press, 2001.

[131] A. Vedaldi, G. Guidi, and S. Soatto, "Moving forward in structure from motion," *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–7, 17-22 June 2007.

[132] B. Leibe, N. Cornelis, K. Cornelis, and L. V. Gool, "Dynamic 3D scene analysis from a moving vehicle," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (Minneapolis, MN), pp. 1–8, June 2007.

[133] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 59, pp. 207–232, September–October 2004.

[134] F. Hoffmann, T. Nierobisch, T. Seyffarth, and G. Rudolph, "Visual servoing with moments of sift features," *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 5, pp. 4262–4267, 8-11 Oct. 2006.

[135] K. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter, "Camera calibration toolbox for matlab." http://www.vision.caltech.edu/bouguetj/calib_doc.

Appendix A Appendix: Additional system model derivations

This appendix contains the derivations of various formulae and analytical expressions given within the body of this work. This Appendix is provided for those interested readers.

A.1 Fitts Law Relationship to Second-order Spring-Mass-Damper Model

Below we provide the derivation of the application of Fitts' law to a second-order system model as proposed by Langolf *et al.* [115, 116] and further investigated by Jagacinski etal [114]. The derivation below provides the intermediate steps between equations (4.8) and (4.9).

$$
A - \frac{1}{2}W = A - A\left(\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\right)
$$

$$
\frac{1}{2}W = A\left(\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\right)
$$

$$
\frac{W}{2A} = \left(\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\right)
$$

$$
\left(\sqrt{1-\zeta^2}\right)\frac{W}{2A} = e^{-\zeta\omega_n t}
$$

$$
\ln\left(\sqrt{1-\zeta^2}\right) + \ln\left(\frac{W}{2A}\right) = -\zeta\omega_n t
$$

$$
\frac{1}{-\zeta\omega_n}\ln\left(\sqrt{1-\zeta^2}\right) + \frac{1}{\zeta\omega_n}\ln\left(\frac{2A}{W}\right) = t
$$

$$
\frac{1}{-\zeta\omega_n}\ln\left(\sqrt{1-\zeta^2}\right) + \frac{\ln 2}{\zeta\omega_n}\log_2\left(\frac{2A}{W}\right) = t
$$

By letting $a = \frac{1}{-\zeta\omega_n}\ln(\sqrt{1-\zeta^2})$ and $b = \frac{\ln 2}{\zeta\omega_n}$, we get the traditional Fitts' Law form

$$t = a + b\log_2\left(\frac{2A}{W}\right)$$

A.2 DES-plant Event Tables and DES-controller State Tables

This section includes the remaining details of DES-plant event and DES-controller symbol definintions. The DES-plant event symbols are given in Table A.1. The description of hypersurface and camera motion combination triggering the event are partitioned by axial component. The first block of symbols describes horizontal motion events. The second block describes vertical motion events. The third block describes depth motion events.

The DES-controller state symbol definitions for the 3-dimensional l_2-model are provided in Table A.2. There are eleven symbols: nine of which relate to some required tracking motion; one for On-Target, and one indicating loss of the target

The DES-controller output symbol definitions for the 3-dimensional l_2-model are provided in Table A.3. The symbols are mapped from the DES-controller state through the DES-controller output function, $\phi(\tilde{s})$. Due to the horizontal axis priority scheme, there $\tilde{r}_2$ and $\tilde{r}_3$ have multiple mappings through the output function.

Table A.1: DES Plant Event $(\tilde{X})$ symbol table for the 3-dimensional l_2-model

Symbol	Description of Plant Event Symbols	Hypersurface	Camera Movement
$\tilde{x}_1$	`On-Target` in horizontal $(x_1(t))$	$h_1(\mathbf{x}) > 0$	left
$\tilde{x}_2$	`Off-Target` in horizontal $(x_1(t))$	$h_2(\mathbf{x}) > 0$	right
$\tilde{x}_3$	`On-Target` in horizontal $(x_1(t))$	$h_3(\mathbf{x}) > 0$	right
$\tilde{x}_4$	`Off-Target` in horizontal $(x_1(t))$	$h_4(\mathbf{x}) > 0$	left
$\tilde{x}_5$	`On-Target` in horizontal $(x_1(t))$	$h_1(\mathbf{x}) > 0$	left
$\tilde{x}_6$	`Off-Target` in horizontal $(x_1(t))$	$h_2(\mathbf{x}) > 0$	right
$\tilde{x}_7$	`On-Target` in horizontal $(x_1(t))$	$h_3(\mathbf{x}) > 0$	right
$\tilde{x}_8$	`Off-Target` in horizontal $(x_1(t))$	$h_4(\mathbf{x}) > 0$	left
$\tilde{x}_9$	`On-Target` in horizontal $(x_1(t))$	$h_1(\mathbf{x}) > 0$	left
$\tilde{x}_{10}$	`Off-Target` in horizontal $(x_1(t))$	$h_2(\mathbf{x}) > 0$	right
$\tilde{x}_{11}$	`On-Target` in horizontal $(x_1(t))$	$h_3(\mathbf{x}) > 0$	right
$\tilde{x}_{12}$	`Off-Target` in horizontal $(x_1(t))$	$h_4(\mathbf{x}) > 0$	left
ε_i	silent events	$h_i(\mathbf{x}) < 0, i = 1 \ldots 4$	
$\tilde{x}_{13}$	`On-Target` in vertical $(x_2(t))$	$h_5(\mathbf{x}) > 0$	down
$\tilde{x}_{14}$	`Off-Target` in vertical $(x_2(t))$	$h_6(\mathbf{x}) > 0$	up
$\tilde{x}_{15}$	`On-Target` in vertical $(x_2(t))$	$h_7(\mathbf{x}) > 0$	up
$\tilde{x}_{16}$	`Off-Target` in vertical $(x_2(t))$	$h_8(\mathbf{x}) > 0$	down
$\tilde{x}_{17}$	`On-Target` in vertical $(x_2(t))$	$h_5(\mathbf{x}) > 0$	down
$\tilde{x}_{18}$	`Off-Target` in vertical $(x_2(t))$	$h_6(\mathbf{x}) > 0$	up
$\tilde{x}_{19}$	`On-Target` in vertical $(x_2(t))$	$h_7(\mathbf{x}) > 0$	up
$\tilde{x}_{20}$	`Off-Target` in vertical $(x_2(t))$	$h_8(\mathbf{x}) > 0$	down
$\tilde{x}_{21}$	`On-Target` in vertical $(x_2(t))$	$h_5(\mathbf{x}) > 0$	down
$\tilde{x}_{22}$	`Off-Target` in vertical $(x_2(t))$	$h_6(\mathbf{x}) > 0$	up
$\tilde{x}_{23}$	`On-Target` in vertical $(x_2(t))$	$h_7(\mathbf{x}) > 0$	up
$\tilde{x}_{24}$	`Off-Target` in vertical $(x_2(t))$	$h_8(\mathbf{x}) > 0$	down
ε_i	silent events	$h_i(\mathbf{x}) < 0, i = 5 \ldots 8$	
$\tilde{x}_{25}$	`On-Target` in depth $(x_3(t))$	$h_9(\mathbf{x}) > 0$	forward
ε_i	silent events	$h_9(\mathbf{x}) < 0$	

Table A.2: DES-controller state ($\tilde{S}$) symbol table for the 3-dimensional l_2-model

Symbol	Description of Controller State Symbols
$\tilde{s}_1$	Off-Target-Depth
$\tilde{s}_2$	Off-Target-Left
$\tilde{s}_3$	Off-Target-Right
$\tilde{s}_4$	Off-Target-Below
$\tilde{s}_5$	Off-Target-Above
$\tilde{s}_6$	Off-Target-Left-Below (horizontal+vertical)
$\tilde{s}_7$	Off-Target-Right-Below (horizontal+vertical)
$\tilde{s}_8$	Off-Target-Left-Above (horizontal+vertical)
$\tilde{s}_9$	Off-Target-Right-Above (horizontal+vertical)
$\tilde{s}_{10}$	On-Target(horizontal+veritcal+depth)
$\tilde{s}_{11}$	No-Target

Table A.3: DES-controller output ($\tilde{R}$) symbol table for the 3-dimensional l_2-model

Symbol	Description of Controller Output Symbols
$\phi(\tilde{s}_1) = \tilde{r}_1$	Move-Forward
$\phi(\tilde{s}_2) = \tilde{r}_2$	Move-Left
$\phi(\tilde{s}_3) = \tilde{r}_3$	Move-Right
$\phi(\tilde{s}_4) = \tilde{r}_4$	Move-Down
$\phi(\tilde{s}_5) = \tilde{r}_5$	Move-Up
$\phi(\tilde{s}_6) = \tilde{r}_2$	
$\phi(\tilde{s}_7) = \tilde{r}_3$	
$\phi(\tilde{s}_8) = \tilde{r}_2$	
$\phi(\tilde{s}_9) = \tilde{r}_3$	
$\phi(\tilde{s}_{10}) = \tilde{r}_6$	Stop: successful acquisition of target
$\phi(\tilde{s}_{11}) = \tilde{r}_7$	Halt: no target present

Curriculum Vitae

NAME:	Duane J. Jacques, B.Sc., B.E.Sc., M.E.Sc.
PLACE OF BIRTH:	Nairobi, Kenya
YEAR OF BIRTH:	1973
POST-SECONDARY EDUCATION	University of Western Ontario 2003 M.E.Sc.
	University of Western Ontario 1998 B.E.Sc. (Computer Engineering)
	University of Western Ontario 1996 B.Sc. (Computer Science)
HONOURS AND AWARDS:	PRECARN Scholars Program 2008: National Graduate Scholarship Faculty of Engineering 2007: Conference Travel Grant Award PRECARN Scholars Program 2006: National Graduate Scholarship Faculty of Engineering 2006: Graduate Thesis Research Award Faculty of Engineering 2005: Conference Travel Grant Award University of Western Ontario 2002/2003: President's Graduate Entrance Scholarship (declined due to deferred enrolment) Sigma Alpha Mu Foundation 2002: Samuel Miller Graduate Scholarship Sigma Alpha Mu Foundation 2001: Samuel Miller Graduate Scholarship Faculty of Engineering 1998: Undergraduate Thesis Project (Semi-Finalist) Certificate of Merit

RELATED EXPERIENCE	Research Associate
	RoadLAB, Department of Computer Science
	University of Western Ontario
	2013 - present

RELATED EXPERIENCE	Lecturer:
	ES1050 Introduction to Design and Innovation
	ECE3375B Microprocessors and Microcomputers
	Department of Electrical and Computer Engineering
	University of Western Ontario
	2008 - 2012

Graduate Teaching and Research Assistant
University of Western Ontario
2000 - 2007

Research Engineer
Mines Research, INCO Ltd.
1999-2001

Publications

`Journal (Refereed):`

D.J.Jacques, K.A.McIsaac, "*An Assistive Device for the Visually Impaired to Perform Guided Reaching Tasks employing a Hybrid Control Scheme*", `IEEE Transactions on Robotics,` *(In Progress)*

D.J.Jacques, K.A.McIsaac, "*Comparison of a Deterministic Finite Automata State Transition Metric to Fitts Law Performance in a Guided Reaching Task*", `IEEE Transactions on System, Man and Cybernetics,` *(In Progress)*

`Technical Reports:`

R.Rodrigo, **D.J.Jacques**, K.A.McIsaac and J.Samarabandu, "*An Object Tracking and Visual Servoing System for the Visually Impaired*", `14th Annual Canadian Conference on Intelligent Systems,` Ottawa, Canada, June 2004.

D.J.Jacques and D. Makrakis, "*MALTA Project: Technical Report*", PRE-CARN Inc., Ottawa, Canada, Sept. 2001.

Dissertations:

D.J.Jacques, "*Modeling and Analysis of Network Traffic for a Teleoperated Load-Haul-Dump Scooptram*", M.E.Sc. Thesis, Faculty of Graduate Studies, University of Western Ontario, London, ON, Canada, Jan.2003.

Conference:

D.J.Jacques and K.A.McIsaac "*Using feature scale to estimate object distance in a reaching task*", `17th Annual Canadian Conference on Intelligent Systems`, Windsor, Canada, May 2008, Poster Presentation.

D.J.Jacques, R.Rodrigo, K.A.McIsaac and J.Samarabandu, "*An application framework for measuring the performance of a visual servo control of a reaching task for the visually impaired*", `IEEE International Conference on System, Man and Cybernetics`, October 2007, Montreal, Canada, p894-901

D.J.Jacques and K.A.McIsaac, "*A Prototype Wearable Assistive Device for the Visually Impaired*", `15th Annual Canadian Conference on Intelligent Systems`, Victoria, Canada, May 2006, Poster Presentation.

D.J.Jacques, R.Rodrigo, K.A.McIsaac and J.Samarabandu, "*An Object Tracking and Visual Servoing System for the Visually Impaired*", `IEEE International Conference on Robotics and Automation`, Barcelona, Spain, April 2005, p3510-3515.

P.Cunningham, T.Charsley and **D.J.Jacques**, "*Development of an Automated 3D Toping System at INCO Limited*", `Proceedings from 5th ISSMA Conference (Telemin1)`, Sudbury, Canada, June 1999, CD-ROM

D.J.Jacques, "*Packet Traffic Behaviour of a MODBUS PLC Application in a Networked Environment*", `Proceedings of the 2002 Electrical & Computer Engineering Graduate Research Symposium`, University of Western Ontario, Editors: T.E.Doyle, W.Greason, Z.Kucerovsky, London, ON, Canada, May 2002, p44-48.

D.J.Jacques and D. Makrakis, "*Modelling and Analysis of Network Traffic Using Self-Similar Processes*", `Proceedings of the 2001 Electrical &`

`Computer Engineering Graduate Research Symposium`, University of Western Ontario, Editors: T.E.Doyle, W.Greason, Z.Kucerovsky, London, ON, Canada, May 2001, p126-128.