UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MOVIT: MONOCULAR VISION-BASED TRACKING

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MUSTAFA A. GHAZI
Norman, Oklahoma
2018

MOVIT: MONOCULAR VISION-BASED TRACKING

A DISSERTATION APPROVED FOR THE
SCHOOL OF AEROSPACE AND MECHANICAL ENGINEERING

BY

Dr. David P. Miller, Chair

Dr. Peter J. Attar

Dr. Andrew H. Fagg

Dr. Mrinal C. Saha

Dr. Zahed Siddique

DEDICATION

To Ammi

Acknowledgments

First of all, I would like to thank my advisor, Dr. David Miller, for all his guidance, support, and motivation throughout this research. I would also like to thank him for his patience while I developed new skills and made new mistakes. I would also like to thank Dr. Andrew Fagg and Dr. Thubi Kolobe for their valuable feedback and advice through various stages of this project.

Then there are those who have directly, or indirectly, enabled me to complete this work: Billy Mays and Greg Williams, for facilitating all my AME machine shop work, Brandt Smith, for enabling all my Fabrication Lab work, Monique Shotande, for helping with my research, Emily North and Amanda Porter, for giving me their perspective on infant movements, Gnana Subramaniam, for helping me with the PNP problem, and Sarah Jamal, for proofreading the final document.

I would also like to thank my family for their love and support over the years. Ammi and Baba have pushed me very hard on this. Murtaza has been a constant source of support. And Naino kindly agreed to move her wedding date to fit my graduation schedule.

Finally, I would like to thank the NSF National Robotics Initiative for funding most of this research.

Contents

List of Figures

List of Tables

Abstract

Cerebral Palsy (CP) is a physical disability that affects approximately 17 million individuals globally. CP can severely impact the development of motor, cognitive, and social skills. Recent research efforts in this domain have led to the development of a series of assistive robot systems designed for crawling-age infants (aged 4-11 months) who are at risk for CP and related motor disorders. These robot systems provide early intervention to mitigate the effects of the above motor disorders. The robot systems capture and interpret infant limb motion in 3D and physically move an infant in response to meaningful crawling-like limb motion. Inertial measurement units (IMUs) are used for the motion capture (mocap) process. IMUs are highly sensitive to electromagnetic fields. Consequently, the presence of electromagnetic interference (EMI) sources in the surroundings causes the assistive robots to malfunction.

Thus the research problem is posed as follows. There is a need for the development of a new mocap approach to replace or augment the existing mocap system for infants. The key requirements are that crawling motions of infants should be captured and the approach must not be sensitive to EMI. The research scope is limited to tracking motion in 3D and does not include methods for automatic gesture recognition or classification. There are two research questions: 1) What are the requirements for capturing crawling motions of infants?

2) To what extent does a mocap system not subject to EMI, meet the above requirements?

The contributions of this research are as follows. Quantitative data on infant crawling motion from past works have been collected and presented in a form useful for the design of mocap systems. A novel approach for mocap based on planar pattern vision markers has been developed. The effects of changing various design parameters on the tracking accuracy has been documented on the basis of physical tests. A performance model has been developed to predict tracking accuracy based on the various design parameters and to allow for comparison with other systems based on tracking planar pattern vision markers.

Key conclusions of this research are as follows. The magnitude of the smallest meaningful crawling motion that an infant can make is 74.6 mm. The worst-case tracking error for the developed system is 19.9 mm. Further evaluation needs to be done to determine whether this is practical for existing gesture recognition and filtering methods.

Chapter 1

Introduction

1.1 Background

The work presented in this dissertation is part of a broader effort to develop a new intervention for medical conditions that result in reduced muscle function and control in infantss. Cerebral Palsy (CP) is a major example of such a medical condition. CP is a life-long physical disability caused by damage to the brain at or around the time of birth. It adversely affects muscle function and postural control. It is the most common physical disability in childhood [29]. According to the Cerebral Palsy International Research Foundation, 17 million individuals around the world have CP [6]. Of these, 50 % live in chronic pain, 33 % are unable to walk, and 20 % are unable to talk. The financial implications are staggering. The US Centers for Disease Control and Prevention has estimated that the cost to care for an individual with CP over their lifetime is nearly $1 million [29].

Unfortunately, there is no known cure for CP. It stays with a person for life. Children with CP learn to walk independently late in life, if at all. There exist

Figure 1.1: SIPPC-3, the latest generation of early intervention robotic devices for infants with CP. Image source: Sooner Magazine/Hugh Scott.

interventions such as physical therapy, medication, and surgery to improve an individual's capabilities. Generally, the earlier the treatment is administered, the better the chances for improvement [14].

There is another cost associated with CP: cognitive development. Crawling is the first form of locomotion available to a child. The most important locomotor experiences are known to be produced by a child's own actions, because self-generated experiences lead to behavior and skill development (Bertenthal et al. [11]). Thus, independent crawling contributes to early cognitive development. For example, an exercise in locating a toy, planning how to get to it, and then physically getting to it helps develop problem-solving and spatial skills. Interaction with other individuals at will helps develop social skills. If a child does not crawl at the appropriate age, then he or she does not experience this exploration phase. As a result, the associated milestones of cognitive development are delayed or even missed.

In conventional medical practice, CP is diagnosed at 18 to 30 months of age,[1]

[1]For conventional screening practices, 9 months is another age for diagnosis but mild cases are less likely to be detected [31].

at which time it is clear that a child has not learned how to crawl and walk on their own (Centers for Disease Control and Prevention [31]). For interventions starting after this age, it can take years before a child can learn to crawl and then walk independently. From a child development perspective, this is a significant delay. The need to mitigate this has spurred research in developing new tests to diagnose CP in younger children. One such research effort has led to the development of a special test for motor disabilities in infants This is the Test of Infant Motor Performance (TIMP) which was developed for use by physical therapists and occupational therapists (Campbell et al. [17], and Campbell et al. [18]). TIMP assesses functional motor performance and is applicable to infants from 32 weeks post conceptional age up to 4 months of corrected age (4 months of age for infants born at term).

1.1.1 A Device for Early Intervention for CP

After the development of a test for early identification of infants at risk of CP [17, 18], the next step was research on early intervention for improvement in their quality of life. The hypothesis[2] is that these infants can benefit from receiving physical assistance during the crawling age. At a minimum, this can enable them to explore the world and continue with their cognitive development. At best, their motor functions can develop at the same pace as that of their typically developing peers. Research in this domain has resulted in a series of experimental assistive devices such as the one described in Pidcoe et al. [69]. To maximize effectiveness, these devices have been designed to be portable enough for home use.

The latest generation of these assistive devices is a robotic mobility device

[2]This is a hypothesis of the intervention and not of this dissertation.

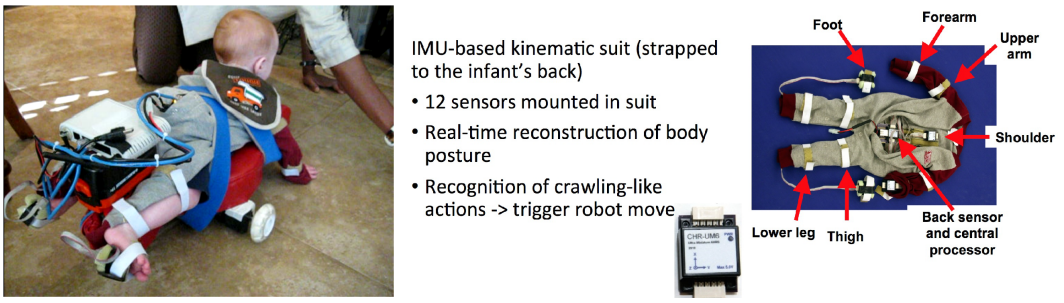


Figure 1.2: Southerland's suit gesture recognition system for the SIPPC robotic devices. It recognizes crawling-like gestures and commands the robot to drive, enabling infants with CP to locomote. The picture on the left shows Southerland's system being used on the SIPPC2, an older generation robotic device. On the right, the details of the suit are outlined. Image source: Southerland [80], Miller et al. [47]

called Self-Initiated Prone Progression Crawler 3 (SIPPC-3), described in Ghazi et al. [35], and shown in Fig. 1.1. Designed primarily for crawling-age infants with CP, it can be used for mobility as well as physical therapy. A major component of its interface is a gesture recognition system by Southerland [80] (see Fig. 1.2). This system captures infant limb motions in 3D and recognizes the crawling-like gestures made. Each individual gesture prompts the robot to drive in a specific short path. Therefore, by making crawling-like gestures, infants can propel themselves along the floor.

The gesture recognition system consists of a "onesie suit" with 12 inertial measurement units (IMUs) sewed into it. The motion capture (commonly referred to as MoCap or mocap) processing is done on a small embedded computer within the "suit." This embedded computer is about the size of a matchbox ($53 \times 36 \times 12$ mm). The gesture recognition processing is done on a separate computer (about the size of a smartphone). The use of these small computers is necessary to make the system portable. The only other things required for the system are a battery for power and a Wi-Fi hub for communications. In

operation, the suit system remains tethered to the battery and Wi-Fi hub by a power cable and a data cable.

1.1.2 Southerland's IMU Suit System

Southerland's IMU suit is a mocap system developed for infants. Mocap systems to capture human motion in 3D are not generally designed for infant applications. Usually, mocap systems designed for adults are adapted for use with infants, e.g. as done by Freedland and Bertenthal [33], Fetters et al. [26], and Wu et al. [93]. Each of these adapted infrared point marker systems that were primarily designed to capture the motion for adults walking upright. Even when adapted, the mocap sessions are constrained. Infants may be restricted to crawl along very specific paths, such as in Freedland and Bertenthal [33]. Alternatively, infants may be immobilized with only the limbs allowed to move freely, such as in Fetters et al. [26], Wu et al. [93], and Olsen et al. [52]. Olsen et al. [52] used a model-based mocap system. In contrast, Southerland's mocap system was designed from the ground up for crawling infants.

Southerland's IMU system was developed for use onboard a mobile robot. This is an important distinction when compared to conventional mocap systems. Generally, mocap systems are not designed for use on mobile platforms. For example, the type of mocap systems used by Freedland et al. [33] and Olsen et al. [52] typically comprise sensors, tripods, computer workstations, and AC power supplies. This can add up to a significant quantity of weight and volume to carry around. Hence, this can make the mocap system much heavier and larger compared to an unencumbered infant. In contrast, Southerland's IMU system requires only the wearable sensors, a Wi-Fi hub and a battery, which

are reasonable payloads for a compact robot carrying a crawling-age baby. The wearable sensors are compact and light enough for an infant to wear comfortably, without being a distraction or a hindrance.

Another novel feature of Southerland's system is the type of gestures involved. Traditionally, limb gesture recognition is focused on the gestures made by individuals who can be upright and make controlled, deliberate movements. For example, the Microsoft Kinect sensor (Zhang [98]) is designed to recognize arm and leg gestures for people standing upright. Breaking away from this trend, Southerland's IMU system was developed to track and recognize exploratory limb motions made by infants learning how to crawl.

1.2 Motivation and Problem Statement

This section outlines the motivating factors for this dissertation and identifies the problem that needs to be solved.

1.2.1 Limitation of Southerland's IMU System

Although Southerland's IMU system [80] is generally reliable, its motion capture process is susceptible to electromagnetic interference (EMI). This is due to the use of microelectromechanical system (MEMS) IMU sensors. A MEMS IMU sensor consists of a MEMS accelerometer, gyroscope, and a magnetometer. Angular rates from the gyroscope are integrated over time to compute 3D orientation in an inertial frame of reference. The accelerometer and magnetometer are used to correct for accumulated error. EMI introduces bias in the magnetometer readings of the MEMS IMUs. This, in turn, introduces errors in the pose estimate of the body.

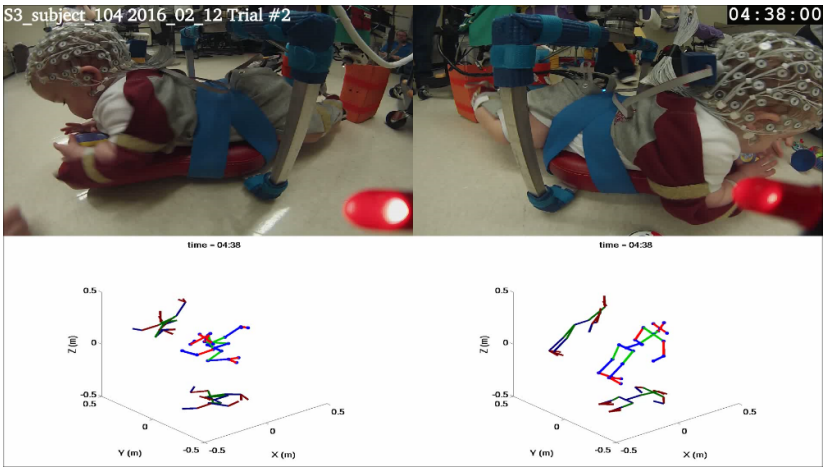


Figure 1.3: Bottom left, kinematic pose estimate as a result of electromagnetic interference (EMI). Bottom right, post-process reconstruction of the actual pose. Image source: Wilson [40].

When pose estimation is inaccurate, the process of interpreting the motion begins to break down. Crawling movements are misinterpreted. The SIPPC-3 robot begins to ignore some crawling movements. It may also respond to the wrong movements. For best results, the SIPPC-3 robot should respond to the desirable crawling movements 100 % of the time. This erratic behavior in the presence of EMI can be confusing for an infant who is trying to learn how to crawl. This limits the utility of the robotic devices in areas with EMI.

EMI can be caused by electrical wiring and ferrous materials in the surroundings. This interference can severely degrade the accuracy of the pose estimation. Fig. 1.3 depicts one such case. Currently, the only solution for dealing with EMI is to avoid areas with power wires and ferrous materials. With this system, there is no reliable way to detect EMI interference other than scanning an area using specialized instruments. Scanning is required because ferrous building structures and electrical wiring are typically hidden under the floor in large buildings so their presence cannot be visually detected. This is not so

much of an issue in homes with wooden structures. But, care must be taken to avoid sources of EMI in homes, e.g., refrigerators and large speakers.

EMI-induced errors in body pose estimation limit the utility of the SIPPC-3 robotic devices. Different locations may have different sources of EMI at different points. To better utilize these robots, the effects of EMI on the motion capture process must be mitigated.

1.2.2 Need for an Alternate Approach

To mitigate the effects of EMI on the mocap process, two approaches are possible. Either Southerland's IMU system [80] can be modified, or a different mocap system can be developed. Wilson [40] developed an approach to improve the body pose estimate from Southerland's IMU system [80]. The approach uses anatomical joint angle limits to apply corrections to the pose estimate. Although Wilson's implementation is offline, it has the potential to be applied online while the motion data are being captured. Essentially, Wilson's work is a step towards modifying Southerland's IMU system. Therefore this dissertation will pursue the other approach, i.e., to develop a different mocap system.

1.3 Problem Definition for Dissertation

The problem that this dissertation seeks to address exists in the context of a series of robotic devices. The context is as follows. Robotic devices were developed to provide early intervention for crawling-age infants (4-11 months). This intervention is intended for use not just in labs and clinics, but also in homes and apartments[3]. Part of the intervention is the 3D motion capture

[3]An apartment can be located within a large building with a steel frame.

(mocap) of crawling-like motions made by the infants. The current mocap system uses MEMS based IMUs strapped onto different parts of the body. The IMUs consist of accelerometers, gyroscopes, and magnetometers. A limitation of this setup is that electromagnetic interference (EMI) results in biased readings from the magnetometers, which in turn introduces error in the mocap system. These errors can be significant enough to cause the gesture recognition process to malfunction, which causes the robotic intervention devices to malfunction. The most common sources of EMI include ferrous metal support structures and electrical wiring under the floor, which are usually hidden from view. Since they are usually hidden, sources of EMI can be difficult to map out and isolate. This limits the utility of this robotic intervention.

In the above context, the problem is defined as follows. A new mocap system must be developed to replace or augment the existing mocap system. A key requirement for the new system is that it must not be susceptible to EMI. In the context of this dissertation, mocap is defined as capturing 3D limb movement. It does not include the process of automatic recognition and classification of those movements.

1.4 Research Questions

The research questions for this dissertation are defined in the context of the development of a new mocap system for capturing infant crawling motions.

1.4.1 Research Question 1

R 1. *What are the requirements for capturing crawling motions of infants?*

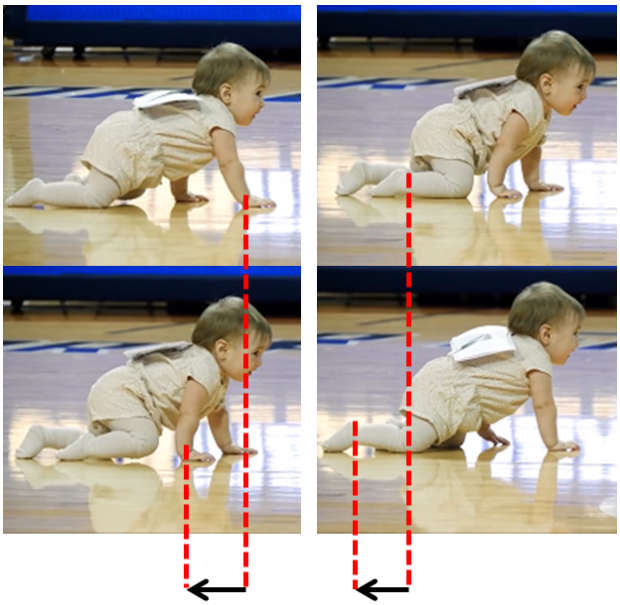


Figure 1.4: Smallest useful crawling motion. The action is somewhat analogous to the stance phase in upright walking. The magnitude of motion is comparable to stride length in upright walking. This motion propels the body. A trained observer is able to detect these from video. A mocap system must be able to detect such motions. Left, arm propulsion. Right, leg propulsion. Source for infant stills: 2016 Rice Baby Race [1].

Motion capture involves tracking points moving in 3D. To develop a new mocap system, design requirements for the motion must be identified. How do the relevant points on an infant's body move when they are making crawling motions? What is the magnitude of the smallest useful crawling motion? Examples of this smallest motion are illustrated in Fig. 1.4. The smallest crawling motion is analogous to the stance phase in upright walking. The magnitude of motion is comparable to stride length in upright walking. This motion is used to propel the body in a desired direction. Such motion can be visually detected by a human observer trained to recognize and classify crawling motion. The accuracy of a mocap system must be sufficient to detect comparable crawling movement. How fast are the motions? Every sensor system has a dynamic

limit which dictates the fastest possible rate of change of a quantity at which it is able to reliably measure that quantity. A mocap system should be able to detect crawling movement even when an infant is moving at its fastest. In this context, the fastest possible crawling movement needs to be identified. What are the sources of data on infant crawling motions? What is the format of the available data? Ideally, trajectories of points should be available in terms of 3D Cartesian coordinates. If trajectories are available in terms of joint angles, then what other data are required to convert those to trajectories in 3D Cartesian coordinates? The crawling development phase usually begins at 4 months of age and typically ends at 11 months. At 11 months, infants' bodies are larger. Are the more stringent mocap design requirements dictated by the smaller 4-month old infants or the larger 11-month old infants?

1.4.2 Research Question 2

R 2. *To what extent does a mocap system, not subject to electromagnetic interference (EMI), meet the above requirements?*

What type of mocap system can be developed such that it is free from EMI? How well can it meet the above mocap requirements? The tracking accuracy must be better than the smallest possible crawling motions (see Fig. 1.4). For example, if the amplitude of the smallest possible crawling motion is x mm, then the mocap system should be able to track within $\pm\frac{x}{2}$ mm. If the mocap system has any dynamic limitations, then the fastest possible crawling motion must be within these limitations. A number of experiments may be required to test the performance against the design requirements. The number of experiments depends on the number of different conditions that can affect performance.

This also raises the question of whether it is feasible to physically test for all the different conditions. It might be more feasible to develop a model to predict performance of the mocap system.

1.5 Research Plan

The research plan is as follows. First, the design requirements for a new motion capture system will be identified. For the application described above, this also includes identifying the kinematics of crawling motions for infants aged 4-11 months. Based on these requirements, a new mocap system will be developed and its performance will be measured. Finally, these performance results will be compared with the kinematics of crawling motions. Evaluation consists of two main criteria. The 3D position tracking accuracy of the system must be better than the smallest crawling motion that an infant can make (detectable by a trained observer, analogous to stride length in upright walking, see Fig. 1.4). The system must be fast enough to capture the fastest crawling motions of an infant.

Chapter 2

Design Requirements: Capturing Infant Kinematics

Prior to developing a new mocap system, the design requirements must be identified. This chapter presents these requirements, both in terms of relevant system features, and in terms of infant kinematics. The emphasis is on kinematics.

In Section 2.1, some desirable system features are identified. These are not necessarily applicable to kinematics, but should still be considered when designing the new mocap system. Section 2.2 outlines some guidelines for deriving kinematic requirements based on crawling motion. It also illustrates why infant size data is important in deriving motion data. Section 2.3 covers the derivation of infant sizes. Section 2.4 derives the infant motion data. Finally, Section 2.5 summarizes the key kinematics requirements necessary for the new mocap system.

2.1 Overview of Desirable System Features

For the application discussed in the previous chapter, the new mocap system must be compact enough to be installed on a mobile robot. Setup and calibration steps should be limited to a few minutes, or it should be possible to perform them in advance. The intervention time with infants using the SIPPC-3 assistive robots is a maximum of 3 consecutive trials with a duration of 5 minutes each[4]. After this time, the infants are likely to disengage, become uncooperative or even distressed.

It is also desirable that the system should not discourage infants from interacting with toys. The act of holding a toy should not compromise the performance of the system. Part of the intervention process is to motivate infants to reach out and play with toys.

2.2 Guidelines for Kinematic Requirements

As discussed in the previous chapter, an infant onboard the SIPPC-3 assistive robot moves its arms and legs. If crawling-like movements are detected, then the robot physically moves the infant in a short path in that direction. The crawling motions for the SIPPC-3 are discussed in detail in Section 2.4.4. A brief description of crawling motion is as follows. For typically developing infants, crawling involves two major components. One is the arm movement for forward or lateral propulsion (see second and third image, Fig 2.1a). The other component is leg movement for forward propulsion (see second and third image, Fig 2.1b). Backward propulsion is not desirable for therapeutic purposes, so it will not be considered. Some crawling movements may be constrained when an

[4]This is the dosage for 1 day. There are 2-3 dosages per week.

(a) Right arm

(b) Right leg

Figure 2.1: Typical crawling cycle for forward propulsion. One and a half cycle is shown. Forward propulsion using the right arm is shown in (a). Forward propulsion using the right leg is shown in (b). For both (a) and (b), the propulsive movement starts at the second image and ends at the third image. Image source: 2016 Rice Baby Race [1].

infant is onboard the SIPPC-3 robot.

The key kinematic requirement is the motion capture system must be able to capture the crawling movements of an infant (the smallest[5] and fastest movements). The crawling motions can be identified by the ankle and wrist movements relative to the hips and upper back, respectively. There are two parts to capturing this motion. First, to detect all the possible crawling motions, the smallest crawling motion serves as the requirement. If the tracking error is sufficient for detecting the smallest possible motion, then it should generally be sufficient for detecting larger motions. The second part of capturing motion is the dynamics. The system must be fast enough to capture motion at all the different body part speeds that are possible. In this case, the fastest possible motion dictates the requirement. If motion can be detected when the body is moving at the fastest possible speed, then motion can also be detected

[5]Detectable by a trained observer, analogous to stride length in walking (Fig. 1.4, 2.1).

at slower speed. Thus, the new mocap system should be able to capture the smallest movements, and should be able to continue detecting movements at fastest speed.

It is desirable that the kinematics data be available in Cartesian space. However, for human motion, kinematics is often reported in terms of joint trajectories. Forward kinematics of the body can be used to convert joint trajectories to 3D Cartesian coordinates. To compute forward kinematics, the sizes of the different body parts need to be identified or derived. Specifically, these are the lengths of the different limb segments.

Using forward kinematics to derive motion data from joint trajectories raises an interesting question. If infant age is not identified for joint trajectories, then what age should be used to select sizing data? Body size increases as an infant grows older. For a given joint trajectory, the older the infant, the greater the range of movement in 3D Cartesian coordinates. Given that the mocap system must be designed for the smallest significant movement (Fig 1.4, 2.1), the sizing data for the youngest infants should be used when computing forward kinematics. On the other hand, if the joint trajectories are to be used for deriving speed, then the sizing data for the oldest infants should be used. This is because for the same joint angle trajectory, a larger body will result in faster speeds for hands and feet.

Infant size data are a prerequisite for computing forward kinematics in case only joint angle trajectories are available. Therefore, these are derived in Section 2.3. After this, the kinematics requirements are identified in Section 2.4.

2.3 Infant Size Data

If motion data are available in terms of joint angles rather than 3D Cartesian coordinates, then those data may be converted to 3D Cartesian coordinates by using forward kinematics. To compute forward kinematics using joint angles, infant sizing data are required. Crawling skills are generally developed between 4-11 months of age. Therefore, this is the age group that will be considered for sizing information.

2.3.1 Survey of Infant Size Data

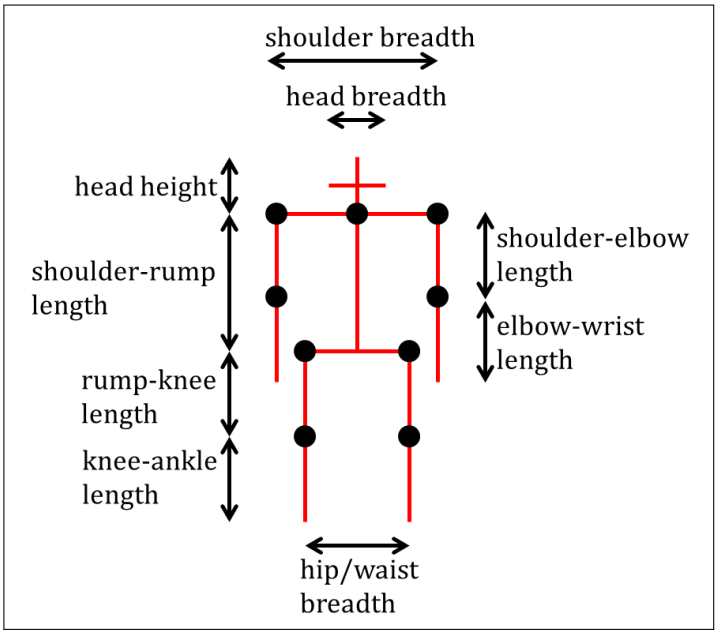


Figure 2.2: Sizing information needed for a simple kinematic model.

Detailed sizing data for 4-11 month old infants do not seem to be well-documented. Details in this context refer to sufficient information to construct a rudimentary kinematic skeleton as illustrated in Fig. 2.2. Child growth standards by the World Health Organization [7] provide only height, arm circum-

ference and head circumference for this age group. Clinical growth charts by the US Centers for Disease Control [30] provide only height and head circumference. Anthropometric reference data by the US Centers for Disease Control [32] provide height, head circumference, and upper arm length. The most detailed source of information seems to be anthropometric data compiled in Snyder et al. [79]. Therefore, these data were used as a reference for infant sizes.

2.3.2 Derivation of Infant Size Data

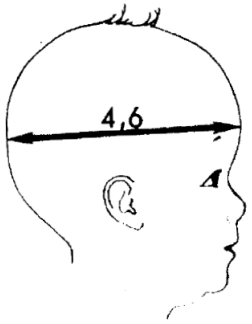


Figure 2.3: Head length measurement. Head height is assumed to be the same as head length. Reproduced from Snyder et al. [79].

For a simplified kinematic skeleton that excludes the hand and feet, 9 measurements are required. These are listed below and are illustrated in Fig. 2.2. Some of them are available in Snyder et al. [79]. The remainder must be derived from other measurements recorded in Snyder et al. [79].

1. head height
2. head breadth (available)
3. shoulder-rump length
4. shoulder breadth (available)

5. shoulder-elbow length (available)

6. elbow-wrist length

7. hip/waist width

8. rump-knee length

9. knee-ankle length

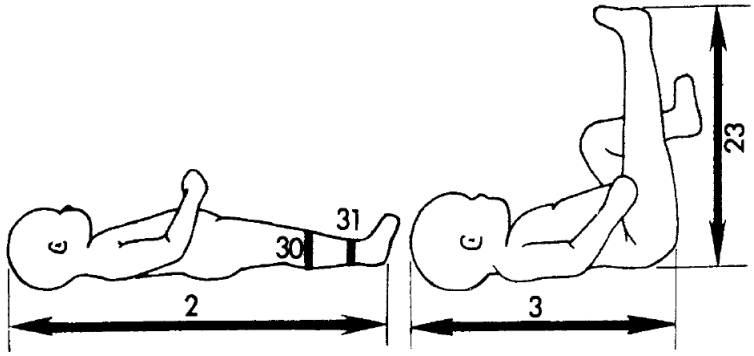


Figure 2.4: Measurements used for deriving shoulder-rump length: crown-sole length (2), crown-rump length (3), and rump-sole length (23). Note that the rump measurements overlap, so the sum is greater than the crown-sole length. Reproduced from Snyder et al. [79].

For lack of any other information, the *head height* is assumed to be the same as the *head length*. The *head length* is illustrated in Fig 2.3.

To derive the *shoulder-rump length*, three measurements are required (see Fig. 2.4): 1) *crown-sole length*, 2) *crown-rump length*, 3) *rump-sole length*.

Ideally, *shoulder-rump length* should be the difference between the *crown-rump length* and *head length*. Unfortunately, the *crown-rump length* and *rump-sole length* are overestimates as seen in Fig. 2.4. Both measurements include all of the rump. This can also be verified by checking the sum of *crown-rump length* and *rump-sole length*. The sum is greater than the height, or *crown-sole length*. This overlapping measurement is given by:

$$\delta_{rump} = \textit{crown-rump length} + \textit{rump-sole length} - \textit{crown-sole length} \tag{2.1}$$

This overlapping meausrement can be divided between *shoulder-rump length* and an *adjusted rump-sole length*. Assuming it is equally divided between the two, we have:

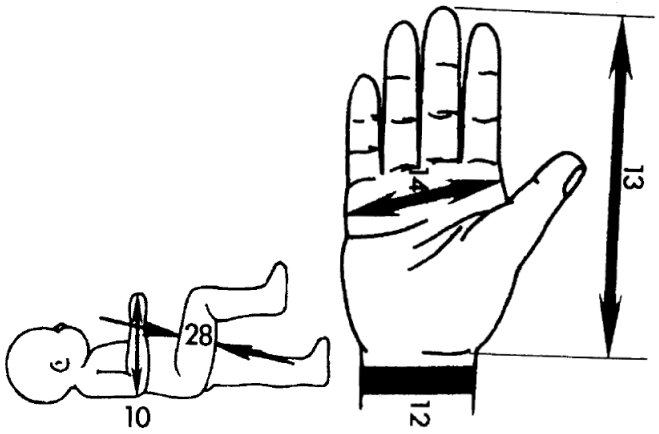


Figure 2.5: Measurements used for deriving elbow-wrist length: elbow-hand length (10) and hand length (13). Reproduced from Snyder et al. [79].

$$\textit{shoulder-rump length} = \textit{crown-rump length} - \textit{head length} - \frac{\delta_{rump}}{2} \tag{2.2}$$

$$\textit{adjusted rump-sole length} = \textit{rump-sole length} - \frac{\delta_{rump}}{2} \tag{2.3}$$

To derive *elbow-wrist length*, *elbow-hand length* and *hand length* are required. These are illustrated in Fig. 2.5. Although this is an overestimate based on the elbow joint, no further measurements are available to deduce the overestimate. Therefore:

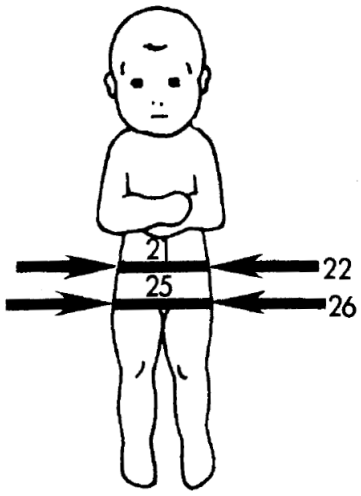


Figure 2.6: Measurements used for deriving the hip/waist breadth: hip breadth (26) and waist breadth (22). Reproduced from Snyder et al. [79].

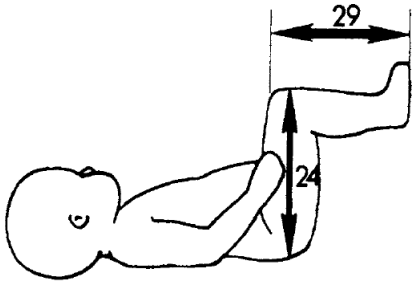


Figure 2.7: Measurements used for deriving the adjusted rump-knee length: rump-knee length (24) and knee-sole length (29). Note that there is an overlap at the rump as well as the knee. Reproduced from Snyder et al. [79].

$$\textit{elbow-wrist length} = \textit{elbow-hand length} - \textit{hand length} \tag{2.4}$$

The *hip/waist breadth* can be based on the larger of the *hip breadth* and *waist breadth* (see Fig. 2.6). The *hip breadth* is always the larger of the two.

The *rump-knee length* is an overestimate based on the rump and knee joints as illustrated in Fig. 2.7. The *adjusted rump-knee length* is given by:

$$adjusted\ rump\text{-}knee\ length = rump\text{-}knee\ length - \frac{\delta_{knee}}{2} - \frac{\delta_{rump}}{2} \tag{2.5}$$

The overestimate δ_{knee}, based on the knee joint, is based on the following equation, where *adjusted rump-sole length* is derived in 2.3 above:

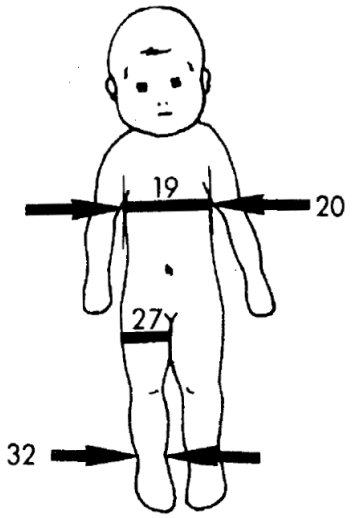


Figure 2.8: Deriving knee-ankle length. Assuming the ankle-sole length is equal to the ankle breadth (32). Reproduced from Snyder et al. [79].

$$\delta_{knee} = (rump\text{-}knee\ length + knee\text{-}sole\ length - \frac{\delta_{rump}}{2}) - adjusted\ rump\text{-}sole\ length \tag{2.6}$$

In terms of the measurements available, this becomes:

$$\delta_{knee} = rump\text{-}knee\ length + knee\text{-}sole\ length - rump\text{-}sole\ length \tag{2.7}$$

The last item is the *knee-ankle length*, which can be derived from the *knee-sole length* (29, Fig. 2.7), *ankle-sole length*, and adjusted for the overestimate

based on the knee joint. Ideally:

$$\textit{adjusted knee-ankle length} = \textit{knee-sole length} - \textit{ankle-sole length} - \frac{\delta_{knee}}{2} \quad (2.8)$$

The *ankle-sole length* measurement is not available. Based on the available measurements, one possible assumption is that it is equal to the *ankle breadth* (Fig. 2.8). Therefore:

$$\textit{adjusted knee-ankle length} = \textit{knee-sole length} - \textit{ankle breadth} - \frac{\delta_{knee}}{2} \quad (2.9)$$

Wrist and ankle diameters can be derived from the wrist circumference (12, Fig. 2.5) and ankle circumference (31, Fig. 2.4). The assumption here is that the wrist and ankle have circular cross-sections.

The resultant sizing data are summarized in Table 2.1. These data will be used to compute forward kinematics where infant limb motion data are available in terms of joint angles.

2.4 Infant Motion Data

The next step is to identify infant motion data. There are two parts to it: speed and range of motion. Speed data are related to dynamic limitations of a mocap system. Range of motion data are related to the 3D position estimation error of a mocap system. Ideally, these data should be available in the form of Cartesian coordinates. Alternatively, if joint angle trajectories are available, then the Cartesian coordinates may be derived by using joint angle data, infant

Table 2.1: Infant sizes derived for forward kinematics. Mean, 5th percentile, and 95th percentile are presented. These data are derived from Snyder et al. [79]. Sizes marked with '*' are direct measurements that did not have to be derived.

	size (mm)								
	3-5 months			**6-8 months**			**9-11 months**		
	$\boldsymbol{\mu}$	**5th**	**95th**	$\boldsymbol{\mu}$	**5th**	**95th**	$\boldsymbol{\mu}$	**5th**	**95th**
head height	146	137	156	155	143	164	160	147	168
head breadth*	114	105	122	118	111	124	122	114	129
shoulder-rump	247	215	263	267	261	280	274	263	274
shoulder breadth*	187	165	204	201	180	220	211	193	231
shoulder-elbow*	123	107	138	131	108	147	145	123	156
elbow-wrist	92	80	102	100	87	109	107	92	119
hip/waist width	143	115	167	159	140	175	166	135	184
rump-knee	101	84	114	111	99	112	124	107	130
knee-ankle	103	99	112	117	109	126	134	114	159
wrist diameter	32	27	35	33	30	36	34	31	38
ankle diameter	37	31	41	39	35	44	41	36	46

sizing data, and computing forward kinematics.

2.4.1 Survey of Motion Data

Data on ankle and wrist movements for 4-11 month old infants do not seem to be readily available in the literature. For this age group, most studies of infant kinematics seem to focus on qualitative results. Freedland and Bertenthal [33] conducted some experiments with crawling infants, but the data they presented were about the cyclic features and patterns of crawling motion. They did not provide quantitative data on limb trajectories. Xiong et al. [95] conducted a study with crawling infants to analyze kinematics and muscle activity. But they, too, published their data in terms of crawling cycles rather than any absolute motions. Righetti et al. [72] published some quantitative data on joint angle

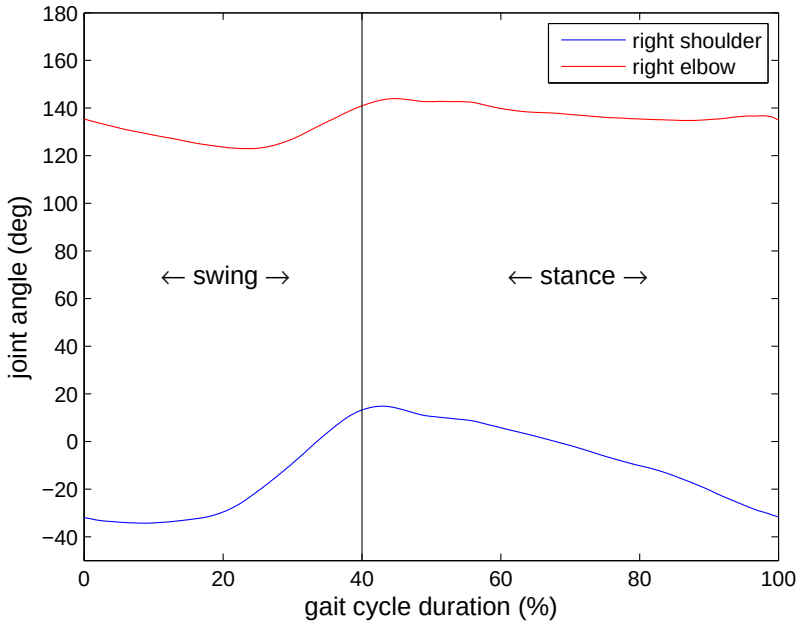


Figure 2.9: Illustration of trajectory data provided by Righetti et al. [72].

trajectories of crawling infants. Other than that, data from Southerland's suit system are available. Kolobe [42] has defined some joint angle movements for crawling infants. These can be detected by trained visual observers.

2.4.2 Speed: Proficient Crawling

Righetti et al. [72] compiled some joint angle trajectories for 7 infants aged 9-11 months. For each infant, they compiled joint angle trajectories for both arms and both legs. Fig. 2.9 illustrates a sample joint angle trajectory of a limb, as provided by Righetti et al. [72]. Note that these data are in terms of joint angles. To derive 3D Cartesian coordinates from the joint angles, forward kinematics using the limb lengths from Table 2.1 must be used. Details on computing forward kinematics are provided in Section 5.3.2.

Righetti et al. [72] presented the trajectories as a percentage of gait cycle

Table 2.2: Infant speeds based on joint trajectories of proficient crawlers from Righetti et al. [72]. Overestimated values are based on the shortest trajectory times. Underestimated values are based on the longest trajectory times. Quoted speeds are 95th percentile.

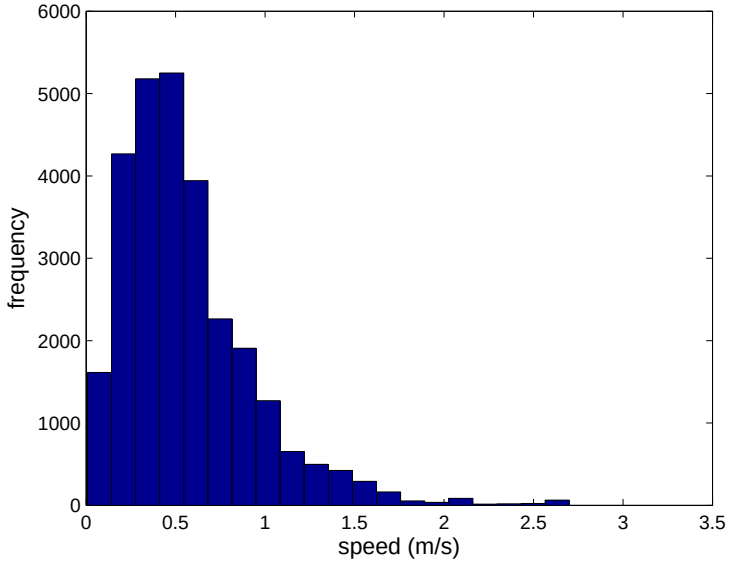
	wrist speed (m/s)	ankle speed (m/s)
overestimate	1.30	0.789
underestimate	0.749	0.434

rather than time. They also quoted shortest and longest median times for the swing and stance phases. To deduce design requirements, the percentage cycle and median time data from Righetti et al. [72] were used to compute the fastest and slowest time series trajectories for all of the joint angles. Infant size data for 9-11 month infants were used to compute Cartesian time series trajectories of the wrist and ankle (using forward kinematics). Speed was computed for each data point in time, and the results are presented in Fig. 2.10 and Fig. 2.11. One is an overestimate (fastest estimate) based on the shortest times, and the other is an underestimate (slowest estimate) based on the longest times. The 95th percentile speeds are presented in Table 2.2.

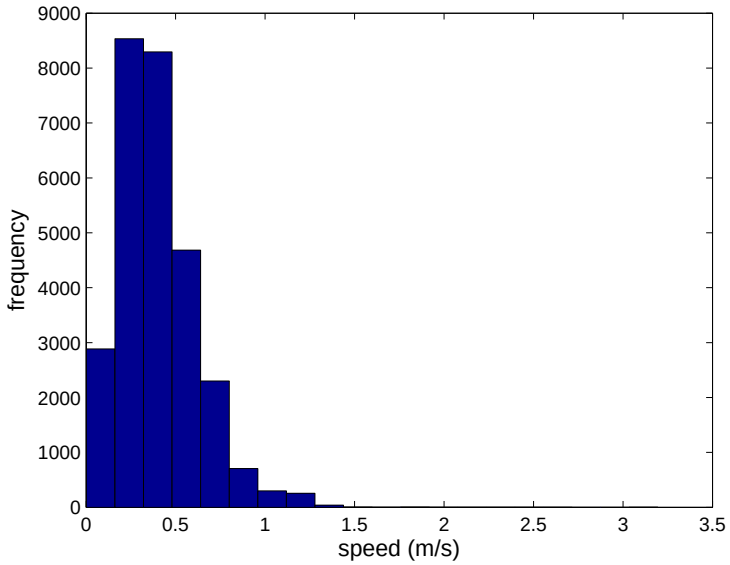
2.4.3 Speed: Infants Onboard Assistive Robot

Wrist and ankle speeds were obtained from motion capture data of typically developing infants onboard the SIPPC-3 assistive robot. These were collected using Southerland's IMU system [80]. Data for 10 infants were used. For each infant, data for 3 sessions were used, resulting in a total sample of 30 sessions for the 10 infants.

The sessions were selected as follows: one at the start of the trial period (week 2), one around the middle of the trial period, and one towards the end

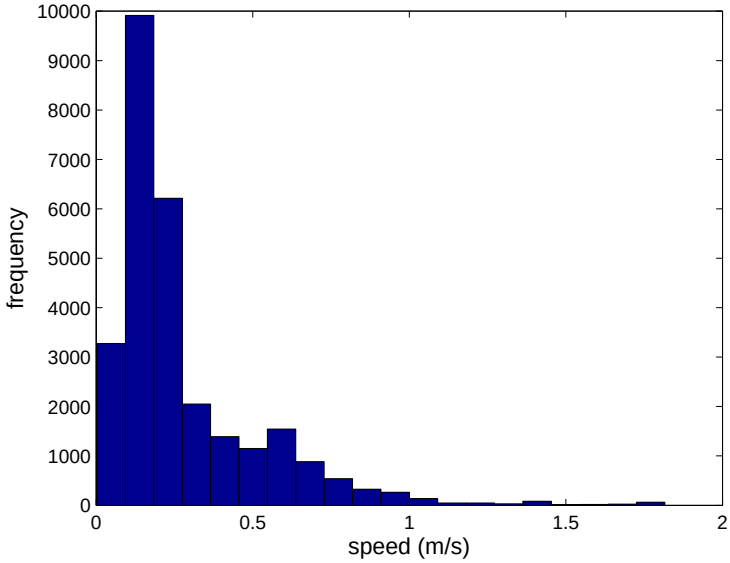


(a) Wrist

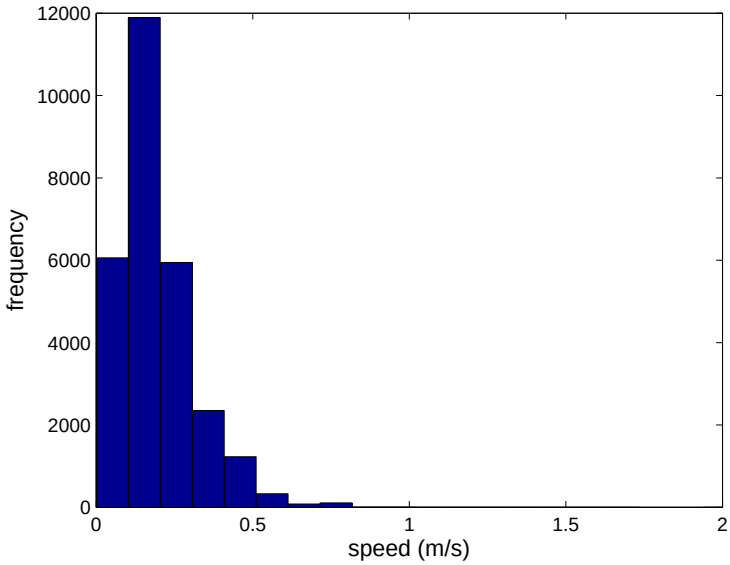


(b) Ankle

Figure 2.10: Histogram of wrist and ankle motion speeds obtained from Righetti et al. [72]. This is an overestimate of speeds because the shortest times were used.

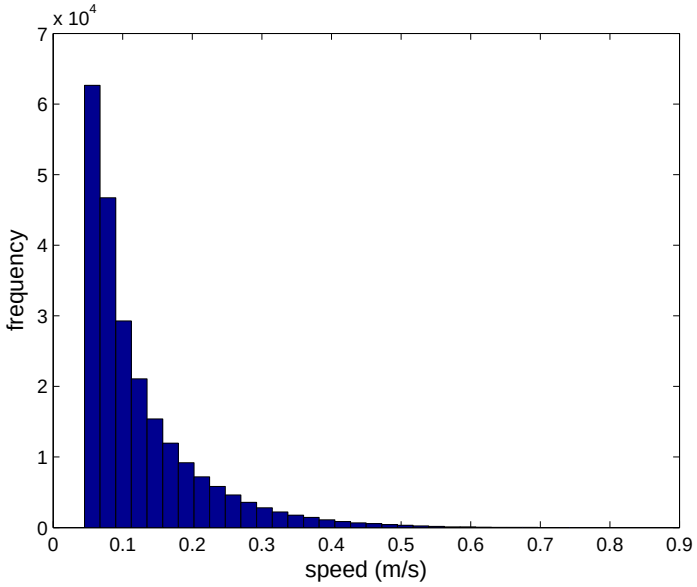


(a) Wrist

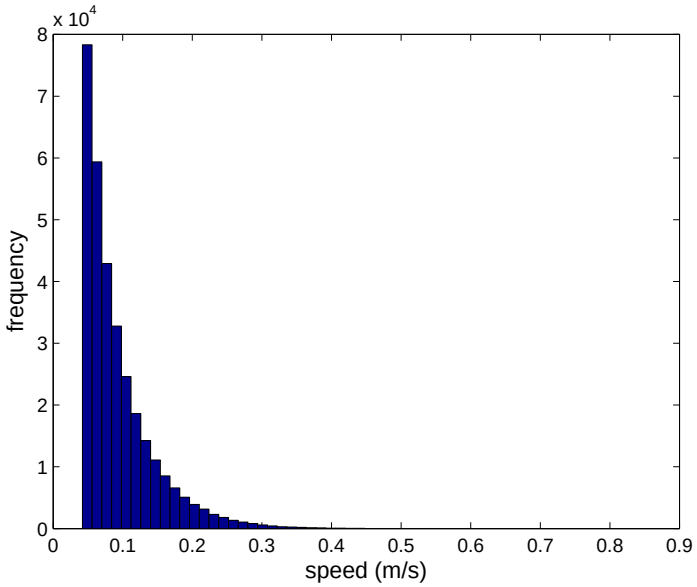


(b) Ankle

Figure 2.11: Histograms of wrist and ankle motion speeds obtained from Righetti et al. [72]. This is an underestimate of speeds because the longest times were used.



(a) Wrist, threshold speed 0.0493 m/s



(b) Ankle, threshold speed 0.0427 m/s

Figure 2.12: Histograms of wrist and ankle motion speeds obtained from the SIPPC-3 robot.

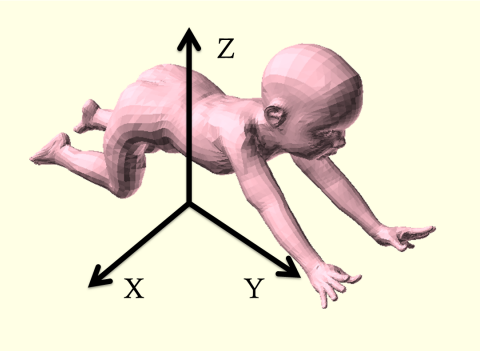


Figure 2.13: Axes used for the figures illustrating range of motion (see Fig. 2.14).

of the trial period. At the start of the trial period, infants had no crawling skills. At the end of the trial period, infants had developed crawling skills. Depending on the pace of development, the middle and end trial week numbers were different for different infants. The shortest middle and end week numbers were 5 and 10 weeks, and the longest were 8 and 16 weeks.

Each session lasted 300 seconds. Data were sampled at 50 Hz, giving 15000 samples per session. Of these, many data points represented no motion. Based on analyses performed by Shotande [77], thresholds of 0.0493 m/s and 0.0427 m/s were used for the wrist and ankle motion respectively. Shotande [77] used a Kolomogrov-Smirnov (KS) distance technique, comparing two distributions, one being of the local maxima speeds and the other of the local minima speeds. This distance computation stems from the KS test, which is a hypothesis test used to determine whether two distributions are similar. The thresholds are the speed at which the distance between the distributions of the maxima and minima is greatest for each limb. These thresholds were used as minimum cut-off values in compiling the speed histograms from the SIPPC-3 data. Essentially, they filtered out data points that did not represent any motion. Inclusion of those data points would have skewed the histograms.

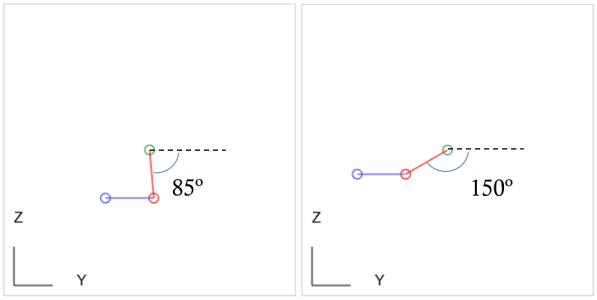
Only peak speeds at these threshold or above were considered to represent motion. Histograms for the wrist and ankle speeds obtained are represented in Fig. 2.12a and Fig. 2.12b. The 95th percentiles for the wrist and ankle were 0.304 m/s and 0.198 m/s respectively.

2.4.4 Range of Motion

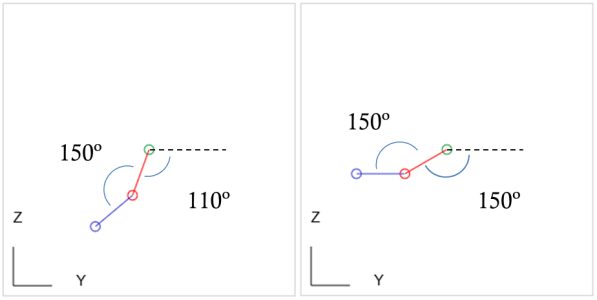
When infants crawl, they propel themselves by using their legs and/or arms. Based on extensive observations of crawling-age infants, Kolobe [42] has defined a set of minimum range of motions. These definitions are used as guidelines for trained human observers to detect crawling motions from videos (e.g. from Fig. 1.4, 2.1). Range of motion is defined in terms of joint angles. It is the minimum movement of one or more joints that is considered to be a deliberate propulsive crawling motion. Given the range of motion in joint angles, the minimum displacement in Cartesian coordinates can be derived by using forward kinematics.

For forward kinematics, the sizing data of the youngest infant age group, i.e., 3-5 months will be used. The rationale for this choice is that the displacement of the wrists and ankles is proportional to limb lengths. The shortest limb lengths will give the smallest displacement, or range of motion, in 3D Cartesian coordinates. When defining mocap system requirements, the smallest range of motion is the worst case. If the tracking accuracy is sufficient to detect the smallest range of motion, then it is also sufficient for the larger motions. For reference, the sizing data are summarized in Table 2.1.

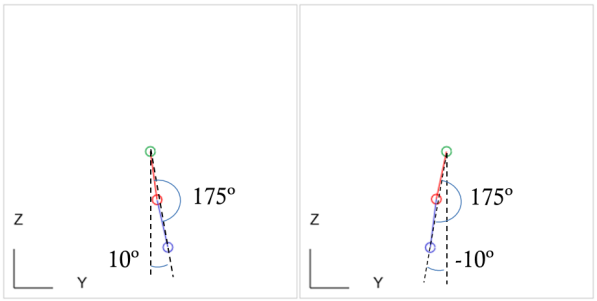
The remainder of this section covers the range of motion for the legs and arms during crawling. For consistency, only the movements of the wrist and



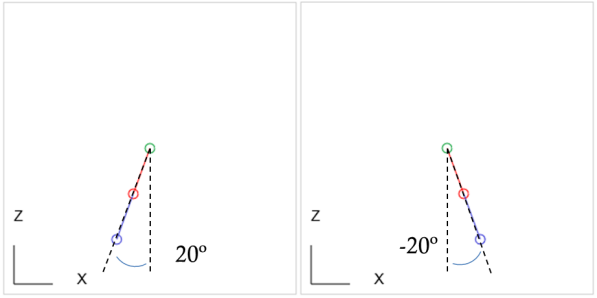
(a) Leg range of motion, unconstrained



(b) Leg range of motion, unconstrained



(c) Arm range of motion, forward propulsion



(d) Arm range of motion, sideways propulsion

Figure 2.14: Leg and arm range of motion. Coordinate frames defined in Fig. 2.13.

ankle joints are considered. These joints are assumed to be points in space, and their position is defined in Cartesian space. The reference axes are illustrated in Fig. 2.13.

Typical forward propulsion using leg motion is defined as follows. Refer to Fig. 2.14a. The hip-thigh angle starts at ≤ 85 degrees and ends at ≥ 150 degrees. The knee-thigh angle changes such that the lower leg is kept approximately horizontal during the motion.

When an infant is onboard an assistive robot, the leg motion is constrained because of the support under the torso (see Fig. 1.2). Therefore, a modified definition of propulsive leg motion is used (see Fig. 2.14b). At the start of the motion, the hip-thigh angle is ≤ 110 degrees, and the knee-thigh angle is 150 degrees. At the end of the motion, the hip-thigh angle is ≥ 150 degrees, and the knee-thigh angle remains constant at 150 degrees.

Forward propulsion using arm motion involves the angle of the wrist with the shoulder joint changing from $\geq +10$ degrees to ≤ -10 degrees. The elbow angle remains constant at 175 degrees. This is illustrated in Fig. 2.14c.

Sideways propulsion using arm motion involves the angle of the wrist with the shoulder joint changing from $\geq +20$ degrees to ≤ -20 degrees. This is illustrated in Fig. 2.14d

From the above, given the same joint angle movements, the smallest significant motions are those made by the infant with the shortest limbs, i.e., the 3-5 months age group. Using the mean limb sizes obtained earlier in this chapter, computed wrist and ankle displacements are provided in Table 2.3. The smallest derived displacement was 74.6 mm.

Table 2.3: Minimum wrist and ankle displacements computed using forward kinematics. Based on range of motion illustrated in Fig. 2.14 and sizing data for the youngest age group of 3-5 months (Table 2.1).

type of motion	**displacement (mm)**
ankle (unconstrained)	108.5
ankle (constrained)	134.8
wrist (forward propulsion)	74.6
wrist (sideways propulsion)	147.1

2.5 Summary of Kinematic Requirements

Based on the compiled and derived results in Sections 2.4.2, 2.4.3, and 2.4.4, a summary of the kinematics requirements is as follows. The magnitude of the smallest possible crawling motion made by an infant is 74.6 mm (wrist movement for forward propulsion). The fastest movement for proficient crawlers is the wrist speed, which can be taken to lie between and 0.749 m/s and 1.30 m/s (based on a single test for each subject at 9-11 months). The fastest possible speed for infants developing crawling skills on the SIPPC-3 robot is also the wrist speed, which is 0.304 m/s (95th percentile, based on 2-3 tests per week for up to 16 weeks). Given that the mocap system to be developed is for infants learning how to crawl, 0.304 m/s is the more appropriate design requirement. When the infants become proficient crawlers and begin moving their limbs at higher speeds, they do not need to be using the SIPPC-3 robot. They should be crawling independently of the robot. All motion must be faster than 0.0427 m/s, which is the threshold for velocity peaks based on ankle movement data from the SIPPC-3 robot.

Therefore, to sum up, the magnitude of the smallest motion to be detected is 74.6 mm. This motion must be detected at speeds ranging from 42.7 mm/s to

304 mm/s. In case the movements exceed the capabilities of the mocap system, ideally, there should be some way of detecting such an event. This does not have to be a requirement but definitely a very desirable feature.

Chapter 3

Developing a New Motion Capture System

From Chapter 1, it follows that an alternative system of motion capture needs to be developed in the context of infant crawling motions. The requirements for capturing these crawling motions were identified in Chapter 2. For the scope of this dissertation, mocap does not include gesture recognition algorithms.

This chapter focuses on the development of a new mocap system and is organized as follows. In Section 3.1, an overview of existing mocap technologies is presented. Section 3.2 covers the selection of the most suitable mocap technology for crawling infants. In Section 3.3, the design of the new mocap system is presented. Finally, a design summary is presented in Section 3.4.

3.1 Types of Systems

Motion capture systems can be divided into two catgories: 1) Inertial systems use IMUs placed on different parts of the body to capture the pose of the

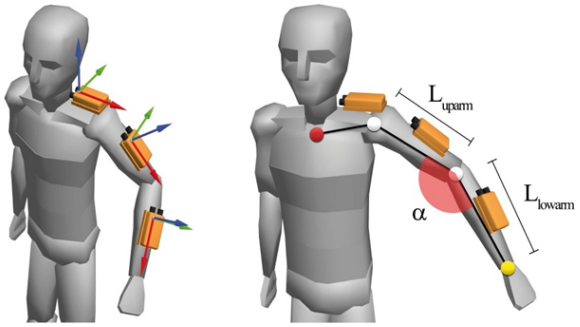


Figure 3.1: Inertial mocap uses IMUs mounted to the body. The IMUs provide orientation which can be used to derive joint angles. Forward kinematics are used to compute full body pose. Image source: [13]

body. 2) Vision-based motion capture systems use cameras to track motion. Although other motion capture systems have been proposed, these two are the most common techniques. Over the years, they have consistently proven to be feasible.

Inertial mocap systems use inertial measurement units (IMUs) placed on the body (see Fig. 3.1). These are almost exclusively based on microelectromechanical systems (MEMS) IMUs. An IMU consists of an accelerometer, a gyroscope, and a magnetometer. Angular rates from the gyroscope are integrated over time to compute 3D orientation (roll, pitch, yaw). The other two sensors are used to correct for drift. MEMS IMUs can be packaged into very small form factors, e.g., the InvenSense MPU-9250 is embedded into a chip that is sized only $3 \times 3 \times 1$ mm [38]. Small form factors of MEMS IMUs allow for relatively unobtrusive placement on the body. A single IMU mounted on a single body part, e.g., the upper arm, measures the 3D orientation of that part. By mounting IMUs to multiple body parts, relative joint angles can be computed (see Fig. 3.1). Given the lengths of the different body parts, forward kinematics can be used to track the position of different parts of the body. The above description is

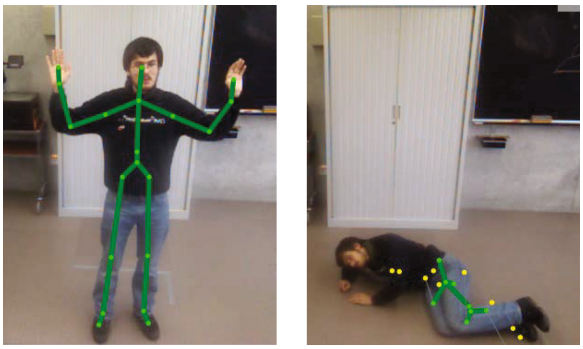


Figure 3.2: Model-based mocap systems capture 3D points and surfaces. Left: predefined body models are used to detect and track the human body. Right: when parts of the body are occluded, detection is unreliable. Image source: Delachaux et al. [22].

greatly simplified, and practical algorithms for IMU mocap are generally much more sophisticated. But, regardless of the implementation, magnetometers are an integral part of an IMU mocap system.

Vision-based motion capture systems can be categorized into 4 distinct types. 1) Model-based systems use cameras to capture the three dimensional shape of the body and infer the underlying skeleton pose. 2) Infrared point-marker systems use cameras to track single-point infrared markers placed on the body. 3) Color marker systems use color-coded markers for tracking. 4) Planar pattern marker systems use specially coded patterns for tracking.

Model-based mocap systems use intelligent algorithms to track the body. Typically, they consist of an infrared projector and an infrared camera (e.g., Microsoft Kinect [98] and Orbbec Astra S [67]). Together, they are used to compute special images called depth images, in which each pixel has an associated 3D location in space. In other words, they capture all the 3D points and surfaces within view of the projector-camera system. Other methods may be used to obtain depth images as well, e.g., the OrganicMotion system [49], which uses multiple color cameras. After a depth image is obtained, clusters

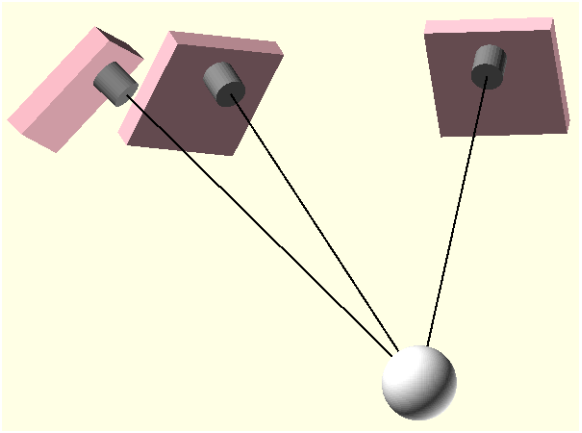


Figure 3.3: The typical approach to tracking infrared point markers or color markers. Multiple cameras are used to track markers in 3D. Each camera provides a line of sight to the marker. The intersection of multiple lines of sight marks the 3D location of the marker.

of pixels representing different parts of the body are identified using predefined body models. Finally, the joint locations are derived from the different body parts (see left, Fig. 3.2).

Model-based mocap systems have some limitations. Since they are vision-based, tracking depends on the body parts being visible to the cameras. If parts of the body disappear from view, then tracking fails. This is known as occlusion[6]. When parts of the body go out of view, the scene no longer contains a complete body which matches predefined body models. Generally, model-based systems assume that a full body, i.e., torso, head, and 4 limbs are in view. If a body part is missing from view, then either the system will fail to detect the body, or it will incorrectly fit the body model to the available depth image (see right, Fig. 3.2). This can happen when one body part is hidden by another body part, e.g., when a person with their back to the camera crosses their arms across their chest. Another issue with matching predefined body models to the depth

[6]Loss of tracking data when the marker or body segment being tracked goes out of view.

image is that clusters of pixels (i.e., objects) connected to the body are considered parts of the body, e.g., a person holding a large object in their hand. A model-based system will either fail to detect the human body because it does not match the predefined body model or it will try to fit a body model by assuming that the large object in the hand is an extension of the body. Both cases are undesirable. Therefore, generally, model-based systems are not reliable when the human body is in contact with other objects or structures. Finally, model-based systems use computationally intensive searching and matching algorithms.

Infrared point marker mocap systems track point markers[7] that are placed on the body. Each marker is associated with a specific point on the body. Hence points on the body are tracked by the use of the markers placed on those points. Markers are made of materials that reflect light very well. Infrared light is flashed at the body and multiple infrared cameras are used to capture the scene. An alternative to reflective markers is to use markers that emit infrared light. Regardless of the approach used, the infrared point markers show up as very bright spots in the infrared camera images. Given the camera pose, a single camera can only be used to compute a line of sight to a marker. Essentially, this is a 3D line originating at the camera and continuing to infinity. To pinpoint the 3D location of the marker, multiple cameras are used. Multiple lines of sight for the same marker intersect at the location of the marker. This concept of using multiple cameras to track a marker is illustrated in Fig. 3.3.

Infrared point marker mocap systems have their limitations, as well. Since they are vision-based, they suffer from occlusion. Tracking depends on the markers being visible to multiple cameras. A single point marker needs to be in the view of at least two cameras to be tracked. If the view of the point marker

[7]These markers are so small that they appear as points in the camera view.

is blocked, then it can not be tracked. Appropriate lighting is also required. Generally, this means that other sources of infrared light are not present and that there are no highly reflective points or surfaces near the markers. Typically, the presence of sunlight interferes with such systems.

Color marker systems use color coded patches for tracking. These are essentially clusters of different colors. Their position can be tracked in 3D in the same manner as infrared point markers, i.e., by using multiple cameras. But they have the additional benefit of the markers being individually identifiable due to their color combination. For example in a marker system with dual colors (two colors for each marker) one marker could be colored red and blue, while a different marker could be colored red and green. Since colored patches are tracked, the markers tend to be a little larger than infrared markers to ensure reliable tracking.

Color marker systems have similar limitations as infrared point marker systems. They suffer from occlusion. If colors are only used for detection, i.e., similar to how infared point markers are used, then the entire marker must be visible to at least two high resolution cameras for 3D tracking. Alternatively, they can be printed as planar shapes with well-defined geometries, in which case a single camera is required for 3D tracking. Appropriate lighting conditions are required. Color marker tracking systems are sensitive to lighting conditions. They must be calibrated for the lighting condition before the tracking process. If the lighting condition changes during tracking, then reliable tracking of colors can become an issue.

Planar pattern marker systems use unique patterns for tracking. Detection is done with the help of grayscale images, so color does not matter, as long as there are two sharply contrasting colors. Usually they are black and white. The

Table 3.1: Motion capture (mocap) technologies that have been proven to be practical.

Type	**Working Principle**	**Limitations**
inertial measurement unit (IMU)	sensor elements that detect motion directly, placed on the body, e.g., [96] [80]	prone to sensor drift and electromagnetic interference (EMI)
model-based, vision	camera, detecting 3D surfaces and using a predefined body model, e.g., [98], [49]	requires appropriate lighting conditions, suffers from occlusion, needs isolated body view, computationally intensive
infrared point marker, vision	multiple cameras, infrared markers placed on the body, e.g., [8], [21]	requires appropriate lighting conditions, suffers from occlusion
color marker, vision	single or multiple cameras, track uniquely colored planar geometries or patches	requires appropriate lighting, suffers from occlusion, color detection sensitive to lighting
planar pattern marker, vision	single camera, track unique grayscale patterns	requires appropriate lighting, suffers from occlusion, detection sensitive to blurring

patterns are printed on planar surfaces. Since different patterns can be defined and detected, individual markers can be uniquely identified. Tracking in 3D is slightly different from infrared point marker systems. Since the markers have well-defined geometries, a single camera can be used to compute the 3D position and orientation of a planar pattern marker.

The limitations of planar pattern markers are similar to the limitations of the other marker systems. Appropriate lighting conditions are required. Occlusion is an issue, though somewhat less since only a single camera is required for

tracking. Motion blur is an issue because if the blurring effects are significant, then the pattern can not be detected reliably. Finally, the marker must be placed on a planar surface for best results.

Table 3.1 summarizes the mocap technologies described above, along with their limitations. For details on these mocap technologies, see Zhou and Hu [99]. As discussed previously, inertial mocap technology has been used in the Southerland's suit system. It has been problematic with its tendency to drift and its susceptibility to EMI.

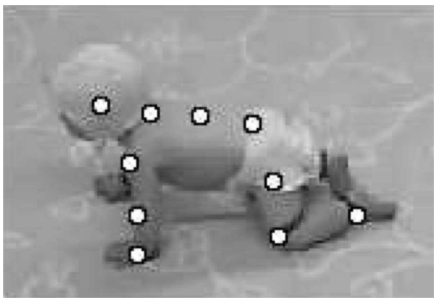
3.1.1 Survey of Mocap Systems Used for Infants

Limited work has been done for capturing limb movements of crawling infants (see Fig. 3.4). Typically, infant motion has been captured by adapting motion capture systems that were developed primarily for adults. Freedland and Bertenthal [33] analyzed changes in crawling limb movement patterns with development (see Fig. 3.4a). Observations began at the age when infants first moved in the prone[8] position (mean age 33.5 $\pm$ 2.5 weeks). Xiong et al. [95] studied muscle activation between upper and lower limbs in crawling infants (ages 8-14 months). Righetti et al. [72] compared crawling movements of infants with quadruped mammals (see Fig. 3.4b). The infant ages were 9-11 months. All three of the above used infrared point marker motion capture systems to record limb kinematics of crawling infants.

In comparison to crawling motion, more work has been done on analyzing motion in other stages of child development. Again, there has been a trend to adapt mocap systems designed mainly for adults. Jeng et al. [39], conducted a comparison study of kicking movements of pre-term and full-term infants in

[8]Lying face down.

(a) Freedland and Bertenthal [33]



(b) Righetti et al. [72]

(c) Meinecke et al. [45]

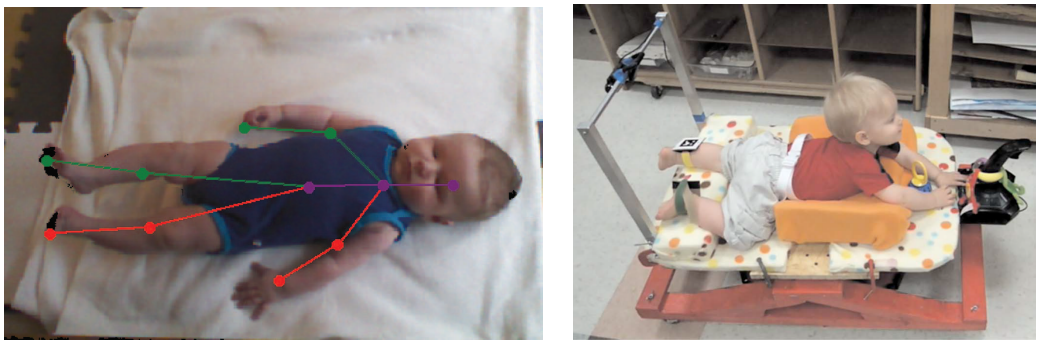
(d) Smith et al. [78]

Figure 3.4: Some examples of mocap systems designed for adults adapted for capturing infant limb motion.

the supine[9] position. Infants were in the age range of 2-4 months. They used an infrared point marker motion capture system. Meinecke et al. [45] studied head, trunk, and leg movements of pre-term and full-term infants in the supine position (see Fig. 3.4c). Infant ages were up to 4 weeks. They used an infrared point marker system to capture movement. Fetters et al. [26] performed a comparison study of kicking movements of pre-term and full-term infants in the supine position. All infants in the study were 5 months old. They used an infrared point marker system to capture motion for this study. Rocha et al. [73] studied reaching movements of infants sitting in a baby chair. Infant ages were 38-41 weeks. Movement data were collected by using an infrared point marker mocap system. Smith et al. [78] analyzed leg movements of infants before they learned how to walk independently (see Fig. 3.4d). Starting ages of the infants ranged from 1-8 months. An inertial system was used to capture the motions. Wu et al. [93] studied arm, leg, and head movements of infants as they made reaching attempts in the supine position. Motion was recorded using an infrared point marker mocap system.

In contrast to the above approaches, some have developed new systems specifically for capturing infant motions. Olsen et al. [51] developed a model-based motion capture system for infants in the supine position (see Fig. 3.5a). Chen et al. [19] developed a planar pattern marker system for infant foot movements in the prone position (see Fig. 3.5b). Finally, as discussed in Chapter 1, Southerland [80] developed an inertial system to capture crawling motions of infants, i.e., in the prone position (see Fig. 1.2).

[9]Lying face up.



(a) Olsen et al. [51] (b) Chen et al. [19]

Figure 3.5: Mocap systems designed specifically for infants.

3.2 Narrowing Selection Pool: Vision Marker

The previous section discussed the types of motion capture technologies available, i.e., inertial, model-based vision, infrared point marker vision, color marker vision, and planar pattern marker vision. This section marks the start of the design process of the new system by selecting the type of motion capture technology to be used.

Inertial systems and model-based vision systems can both be ruled out. The objective of this work is to complement the limitations of an inertial motion capture system. Therefore, an inertial system will not be considered for the new design. Model-based vision systems search for predefined body shapes and poses, so parts of the body going out of view is an issue (see right, Fig. 3.2). Furthermore, commercially available systems do not appear to include models of crawling infants. Even if a custom system based on crawling infant body shapes is developed, infant interaction with toys would cause errors and even failure. This is because a model-based system detects 3D surfaces in space. Connected surfaces will generally appear to be part of the same body. Different bodies in the scene are compared against predefined human body models. Without

information about the geometry of every single object to be touched by an infant, it is challenging to distinguish between an infant's hand and a toy, in real-time. The toy will appear as an extension of the infant's body. Either the predefined human body model will be incorrectly matched to this apparently abnormal body, or the body will not be detected at all. This issue will be compounded if the infant's hand should touch any part of the SIPPC-3 robot: a large structure that is within reach of the infant (see Fig. 1.1). Although Olsen et al. [51] have developed a model-based system for infant motion capture, in their case, the only 3D body visible to the system was that of the infant to be tracked (see Fig. 3.5a). There were no other bodies, such as toys or robot structures, that could interfere with detection. Therefore, the remaining choices are infrared point marker vision, color marker vision, and planar pattern marker vision systems.

3.2.1 Initial Selection: Color Marker Vision

As mentioned above, the choices have been narrowed down to marker-based vision approaches (infrared point, color, and planar pattern). The next step is to identify the specific approach to use. By far the most popular type of system is the infrared point marker vision system (e.g., the Vicon system [8]). A large selection of such systems is commercially available (see Table 3.2). An infrared point marker vision system can be somewhat limiting in the context of capturing infant crawling motion of infants, where markers may have to be applied and calibrated within minutes, and where the entire system must fit onto a portable robot not much larger than an infant. A significant amount of skill and time is required for infrared point marker application and system calibration. Infrared

point marker systems are also unsuitable for use on a mobile robot because they operate on the assumption that the cameras are fixed relative to one another. While this is a reasonable assumption for cameras mounted to distant external structures, it cannot be assumed for cameras mounted on multiple mobile robot masts. These cameras can be knocked around, either during transport, or by infants during operation.

In an attempt to find an infrared point marker system that does not require bulky support equipment, a number of commercially available systems were surveyed. Table 3.2 lists the surveyed infrared point marker camera systems. With the exception of the OptiTrack Slim 3U [64] and Advanced Realtime Tracking SMARTTRACK [4], all systems require bulky support electronics. The OptiTrack Slim 3U camera has some onboard vision processing and has development tools available, apparently for use with any computer running Windows OS. This opens up the potential for use with a portable, small form factor, embedded computer using Windows OS. However, it still has the other limitations described above, i.e., it requires a significant amount of time and skill to apply the markers and calibrate the system. It also assumes that cameras remain undisturbed once the system has been calibrated. The Advanced Realtime Tracking SMARTTRACK camera system [4] is somewhat compact and portable. Two cameras along with support electronics are all built into a single housing of size 410 $\times$ 90 $\times$ 60 mm. Its limitations are that it is not scalable to multiple units and is limited to tracking a maximum of 4 markers.

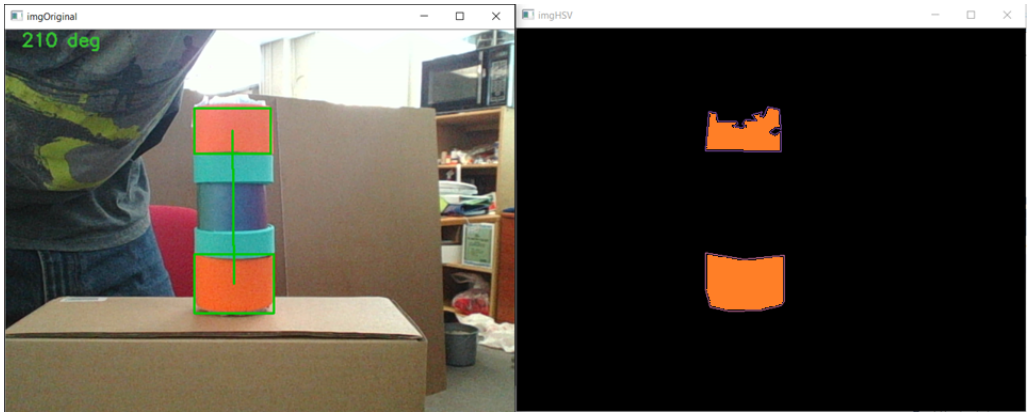
While infrared point marker systems dominate the commercially available mocap systems, these are not the only type of marker that can be used for tracking. As mentioned in Section 3.1, other types of markers include colored markers and planar patterns. Colored markers have been popular with tracking

Table 3.2: List of commercially available infrared point marker motion capture camera systems. Generally, they are too bulky for use on a portable mobile robot like the SIPPC-3.

Company	**Product**
Vicon	Vantage [91]
	Vero [90]
Qualisys	Miqus [70]
	Oqus [71]
OptiTrack	Prime 41 [62]
	Prime 17W [61]
	Prime 13 [59]
	Prime 13W [60]
	Flex 13 [57]
	Flex 3 [58]
	Slim 13E [63]
	Slim 3U [64]
	V120 Duo [65]
	V120 Trio [66]
PhaseSpace	Impulse X2 [68]
Codamotion	3D Motion Analysis System [21]
Phoenix Technologies Inc.	VisualEyez III [87]
	VZ4050 [89]
	VZ4000v [88]
BTS Bioengineering [12]	SMART-DX 100
	SMART-DX 400
	SMART-DX 700
	SMART-DX 6000
	SMART-DX 7000
Advanced Realtime Tracking	ARTTRACK5 [2]
	ARTTRACK5/C [3]
	TRACKPACK/E [5]
	SMARTTRACK [4]
Xcitex	ProCapture [94]
InnoVision Systems Inc.	MaxPRO 3D System [83]
	MaxTRAQ 3D System 60/120 fps [85]
	MaxTRAQ 3D System 160/500 fps [84]
	Max100 Mocap System [81]
	Max300 Mocap System [82]

(a) Wang et al. [92]

(b)



(c)

Figure 3.6: Mocap with color markers: (a) shirt with colored patches by Wang et al. [92], (b) initial marker design using dual-color bands at each limb, (c) tracking orange color from the bands.

applications. Planar patterns have been used primarily for detecting pose in augmented reality applications. Detecting colored markers is relatively less computationally intensive than tracking planar markers, which require pattern detection algorithms.

The sole commercially available color marker system seems to be the CMUcam5 Pixy [20]. This consists of a camera with an onboard processor that can be calibrated to detect multicolor markers (uniquely identifiable clusters of colors). This is a detection system only. It detects markers in the camera image.

Several applications have used colored markers but under constrained conditions. Sargent et al. [75] described a system for tracking soccer playing microrobots; this system used a fixed camera, planar color markers on each robot and stable lighting covering the entire field. Miller et al. [48] used similar tracking hardware to compute position and orientation using a 3D colored marker in the domain of spacecraft docking. Miller et al. [48] assumed smoothly changing lighting. Breitenmoser et al. [16] used a colored marker system for robot localization. This system was also designed for fixed background color and lighting conditions. A colored marker system to be tracked on a cell phone was developed by Bagherinia et al. [10]. The underlying assumption was that the colors on the markers are uniformly illuminated, i.e., with no shadows on the colors. It was also assumed that the marker was not rotated by more than 40 degrees from the expected orientation. A system to detect joint locations in color images was implemented by Nergui et al. [50]. They used colored bands at the joints. Meyer [46] developed a color marker tracking system for controlling the yaw and altitude of a toy helicopter. The color marker was a ring divided into 3 segments of 3 different colors. This system was reliable under even lighting conditions.

Wang et al. [92] developed perhaps the most feasible color marker system in the context of tracking human body motion. They developed an upper body motion capture system based on a specially designed shirt with colored patches (see Fig. 3.6a). The system used a geometric model of the body wearing the shirt. The accuracy of the system depended on how similar the wearer's body was to the internal model. Although Wang et al. [92] demonstrated that their system could deal with changing white balance, in practice, they only had to deal with changes due to the subject moving around in the image. The background and, one can argue, a large part of the image, remained constant. In the case of a mobile robot with onboard cameras, the background is constantly changing as the robot drives around, so the white balance can change significantly. A major limitation of this system, with respect to the mobile robot, is of the amount of effort required in reconfiguring the system. If the color pattern on the shirt in Wang et al. [92] is changed slightly, then the precomputed 80,000 color textured body model poses would have to be recomputed.

Given the apparent successes of previous attempts in tracking colored markers, colored markers were initially selected for the new system. A known issue with infrared point markers is occlusion resulting from limb rotation: for infrared point markers placed on a limb, the markers go out of view when the limb is rotated away from the camera. To avoid this issue with color tracking, colored bands were designed instead of individual markers. Fig. 3.6b and 3.6c illustrate this approach. A band worn on the wrist or ankle would be visible from all directions. Nergui et al. [50] used colored bands in their tracking system, most likely for the same reason. To reduce the number of false positives, and to increase the number of uniquely identifiable markers, dual-color bands were used. The number of uniquely identifiable markers can be increased as fol-

lows. If there are n possible colors that can be tracked, then there can only be n uniquely identifiable single-color bands. But for dual-color bands, $n \times (n-1)$ color combinations or uniquely identifiable bands are possible.

3.2.2 Abandoning Color Marker Vision

As mentioned above, the color marker vision approach was initially selected for this research. This decision was motivated by the apparent success and reliability of previous works that used color markers (these works were described in Section 3.2.1). Unfortunately, preliminary testing with color markers demonstrated otherwise. The important role of relatively fixed lighting and white balance in previous works using color markers became apparent. From testing, it was clear that color marker detection is not very robust to changing lighting (brightness and color) and background color conditions. Therefore, the color marker vision approach was abandoned.

Various color spaces[10] were used but color markers could only be tracked under relatively fixed lighting conditions. The color spaces used for testing were RGB, HSV, and CIE Lab. A custom space with 3 channels defined as R-G, G-B, and B-R was also used. This was inspired by the YCrCb color space.

A major challenge in tracking colors consistently under changing lighting conditions is that existing color spaces were developed for colors on planar surfaces, e.g., print media and electronic displays. They were not designed to capture the effects of surface curvature and texture. These effects exacerbate the apparent change in color due to changes in lighting.

[10] A color space is a coordinate system that defines a color.

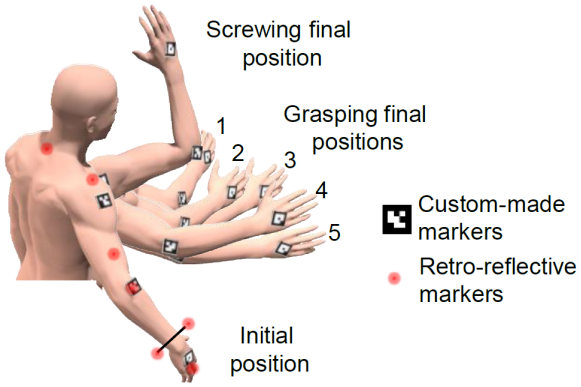


Figure 3.7: Example of planar pattern markers used for mocap. Image source: Bonnet et al. [15].

3.2.3 Revised Selection: Planar Pattern Marker

In contrast to colored markers, planar pattern markers are very robust to changes in illumination, background color, and shadows. Another advantage that they offer is that a single camera can be used to track planar pattern markers in 3D. This includes not just 3D position, but 3D orientation as well. As is the case with color markers, planar pattern markers can be uniquely identified. These features have been discussed in Section 3.1. These markers have been popular in Augmented Reality (AR) applications but have been relatively unheard of in motion capture. Chen et al. [19] and Bonnet et al. [15] are two examples of use of planar markers for human motion capture (see Fig. 3.5b and Fig. 3.7). Bonnet et a. [15] did not use AR techniques to track the markers, however. They used markers only as textures to be tracked in a 2D image and used kinematics to compute the positions of the markers.

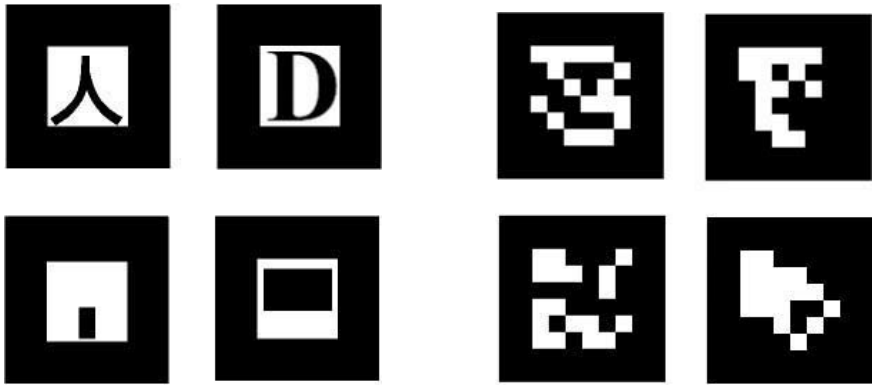
The general working principle to track a planar patterned marker in 3D is as follows: 1) a planar marker with a unique pattern is detected in a gray-scale image, and, 2) the marker position and orientation are computed by using control points on the marker image, knowledge of the geometry of those points,

and applying principles of projective geometry. A limitation of this approach is that marker patterns are being detected, and there is a limit to how far patterns can be from the camera and still be recognized. A marker must cover sufficient pixels in the image for the pattern to be resolved reliably.

A new motion capture system based on planar pattern markers will be developed. A number of planar pattern markers have been developed for augmented reality (AR) applications over the years, e.g. ARToolKit [41], ARTag [27], and AprilTag [53]. The new Monocular Vision-Based Tracking (MoViT) motion capture system will be based on a relatively recent planar marker system called ArUco [34]. ArUco markers have error correction built into their pattern. The ArUco system defines fewer permutations of marker codes than are mathematically possible in order to maximize the distinction between the different marker codes. ArUco marker tracking is available as a contributed module for the Open Source Computer Vision (OpenCV) library version 3.1.0 [55].

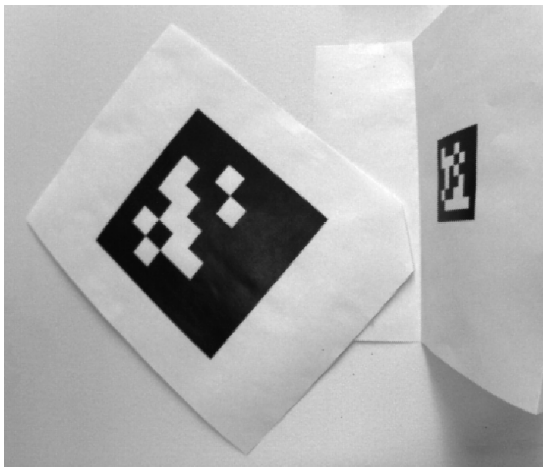
3.3 System Design

In the previous section, a planar pattern marker-based approach was selected for the new motion capture system called MoViT (Monocular Vision-Based Tracking). The MoViT system will be based on a planar pattern marker called ArUco [34]. Unlike conventional infrared point marker systems which require multiple cameras for tracking, planar pattern markers can be tracked using a single camera. Hence the name. Based on these design choices, this section presents the design of the MoViT motion capture system. First, the locations for the markers on the body are identified. Then a marker bracelet concept and marker border modification (MoViT marker) is presented. This is followed by



(a) ARToolKit [41]

(b) ARTag [27]



(c) AprilTag [53]

(d) ArUco [34]

Figure 3.8: Examples of planar pattern markers. For the new motion capture system, a relatively recent planar pattern marker system called ArUco has been selected. Image source for (a) is Fiala [27].

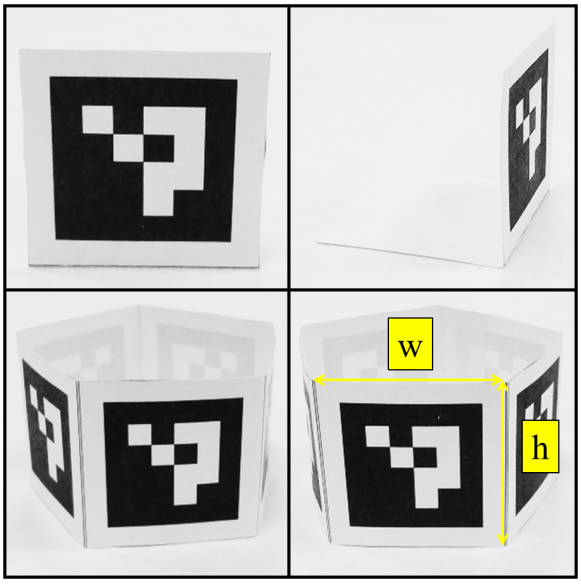


Figure 3.9: Single marker (top) which is useful for augmented reality applications, compared to the marker bracelet design (bottom) which is more practical for motion capture. A single marker (top left) can go out of view when rotated (top right). But with the MoViT bracelet design (bottom left) there is always at least one marker visible after any rotation (bottom right). The height of the bracelet is h and the width of each face is w.

the geometrical design of the marker bracelet. Finally, dynamic design considerations are presented.

3.3.1 Marker Placement on Body

The first step in designing a marker-based motion capture system is to identify where the markers will be placed on the body. As discussed in Chapter 2, the wrist and ankle locations can be used to detect crawling motions. Therefore, the most convenient locations for marker placement are just before the wrist and ankle joints. Reference markers to transform tracked coordinates to the infant body frame of reference can be placed on the back and/or hips. This is illustrated in Fig. 3.10.

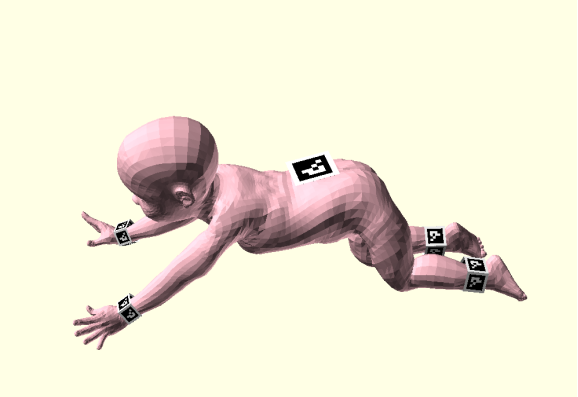


Figure 3.10: Illustration of the locations on which the marker bracelet may be worn by a crawling infant. A reference MoViT marker is attached at the lower back.

The coordinates of a target marker in the infant body frame of reference, i.e., in the frame of reference of the reference marker, are equivalent to the translation vector below:

$$^{REF}\vec{T}_{M/REF} = {}^{REF}R_C.{}^{C}\vec{T}_{M/REF}, \tag{3.1}$$

where $^{REF}\vec{T}_{M/REF}$ is the target marker translation vector relative to the reference marker, defined in the reference marker frame, $^{REF}R_C$ is the rotation of the camera frame relative to the reference marker frame, and $^{C}\vec{T}_{M/REF}$ is the target marker translation vector relative to the reference marker, defined in the camera frame. Typically, in planar pattern marker tracking, $^{C}\vec{T}_{M/REF}$ and $^{REF}R_C$ are not directly available. Rather, all translation vectors and rotations are computed relative to the camera, and in the camera frame of reference. $^{C}\vec{T}_{M/REF}$ can be derived as follows:

$$^{C}\vec{T}_{M/REF} = {}^{C}\vec{T}_{M/C} - {}^{C}\vec{T}_{REF/C}, \tag{3.2}$$

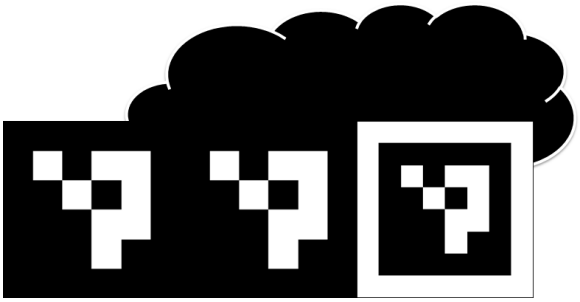


Figure 3.11: Illustration of how the ArUco marker's outline can become indistinguishable from a dark object in the background, or from the outline of an adjacent marker. All four edges are required for detection. To deal with this issue a white border has been added. This is the MoViT marker (right). Note that this reduces the size of the marker that is detected by the ArUco library.

where ${}^{C}\vec{T}_{M/C}$ is the target marker translation vector relative to the camera, defined in the camera frame of reference, and ${}^{C}\vec{T}_{REF/C}$ is the reference marker translation vector relative to the camera, defined in the camera frame of reference. ${}^{REF}R_{C}$ can be derived as follows:

$$ {}^{REF}R_{C} = ({}^{C}R_{REF})^{-1} = ({}^{C}R_{REF})^{T}, \tag{3.3} $$

where ${}^{C}R_{REF}$ is the rotation of the reference marker frame relative to the camera frame. Its inverse is equivalent to ${}^{REF}R_{C}$. If ${}^{C}R_{REF}$ is a rotation matrix, then the inverse $({}^{C}R_{REF})^{-1}$ is equivalent to the transpose $({}^{C}R_{REF})^{T}$.

3.3.2 Bracelet to Increase Visibility

The second step of the design is to determine how to place the markers on the body. A novel marker design in the form of a bracelet is presented here (see Fig. 3.9). As mentioned previously, this is not the first approach to use planar patterned markers for motion capture of infants. Chen et al. [19] used planar patterned markers from the AR system called ARToolKit [41] for capturing

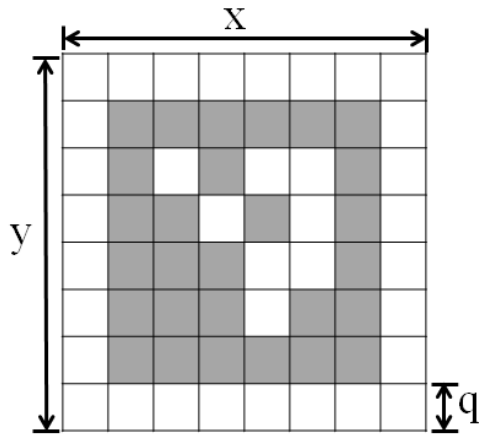


Figure 3.12: The MoViT marker is an 8×8 grid of black and white squares. The outermost squares are white. The squares inside them form a black border. The inner 4×4 grid of black and white squares encode the marker's unique identity. The marker size is s where $s = x = y$. The size of each small square is q.

infant kicking motion. They mounted one marker to each ankle (see Fig. 3.5b). A single marker is sufficient for the back and hips, but not for the wrists or ankles. This is because a marker could disappear from view when rotated away from the camera. Hence the proposed marker bracelet design, with planar markers all around the periphery (Fig. 3.9). If there is any rotation about the axis of the bracelet, then at least one marker will always be visible. This concept is similar to the dual-color marker band approach presented at the end of Section 3.2.1 (Fig. 3.6). Fig. 3.10 illustrates an infant model wearing bracelet markers at the wrists and ankles, with a single reference marker at the lower back.

3.3.3 Contrasting White Marker Border

Traditionally, planar pattern markers for AR applications, including the ArUco marker, have been designed with the expectation that markers will be printed

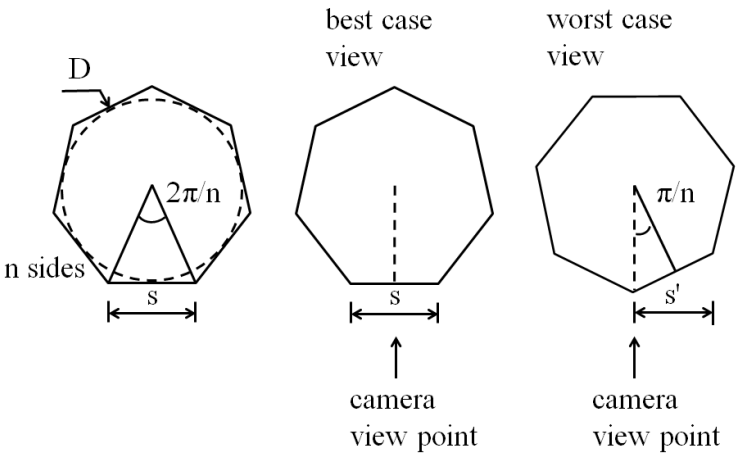


Figure 3.13: Best and worst case geometry views. The MoViT marker geometry is defined by an n-sided polygon with an inscribed circle of diameter D.

on a light background which will provide a contrast against the dark marker borders. Designers of such markers also expect that clusters of markers printed together will have spacing between them. Detection of marker borders with the aid of these color-contrasting edges is a prerequisite for planar pattern marker detection. Since a marker border is defined by four color-contrasting edges, all four edges must be detected in order to successfully detect a marker border.

The prerequisite of detecting a marker border introduces two challenges in using planar pattern markers for the motion capture system. First, the background of the marker may not necessarily be a light color and any marker borders in front of such a background may not be distinguishable. Secondly, adjacent markers on the bracelet will have no distinguishable border between them. The absence of detectable borders in both these cases is illustrated in Fig. 3.11.

To guarantee the marker border detection on all four sides, the ArUco marker has been modified to create the MoViT marker. This has been done by introducing a white border around the original ArUco design (see Fig. 3.11 and

Fig. 3.12). For a fixed area, this reduces the size of the marker detected by ArUco detection software. But it improves the chances of detection for the motion capture application. This improvement is quantified by an experiment in Section 4.1.

The white border has been defined to be as wide as the black border. This way, as long as the black border is visible, the white border that distinguishes it from the surroundings will also be visible. The white border has been incorporated in the illustration of the bracelet design in Fig. 3.9.

3.3.4 Bracelet Geometry

Given the bracelet design and the modification of the marker pattern with the white border, the next step is to determine the bracelet geometry. The markers making up the bracelet should be large enough to track, and the bracelet should not be too bulky. This subsection presents the design process based on these desirables. Bulkiness is defined later in this text.

The geometry of the marker bracelet is defined by an n-sided polygon with an inscribed circle of a diameter D, where D is the diameter of the body part on which it is to be worn (left, Fig. 3.13). Each of the n sides represent a marker face of size $s = w$ (see left, Fig. 3.13, and bottom right, Fig. 3.9). There are three major factors affecting the design. 1) If height = width, $h = w$ (the faces are square) then size s of each side decreases as n increases (assuming D is constant). Ideally, s should be as large as possible to achieve maximum accuracy and minimum motion blurring. 2) The worst-case projection angle of s decreases as n increases. This means that the ability to track the marker face at an angle improves as n increases. 3) The bulkiness factor is related to

the size of the circle that circumscribes the bracelet. Bulkiness of the bracelet decreases as n increases.

To understand the effects of the first two factors, consider the following. Ideally, s should be as large as possible to achieve maximum accuracy. In the best case, the marker is perpendicular to the camera view, and s is maximum when n is minimum (see center, Fig. 3.13). This is the first factor mentioned above. In other cases, the marker is rotated by an angle so that only a projection $\acute{s}$ is visible (see right, Fig. 3.13). The larger the rotation angle, the smaller the projection $\acute{s}$. For any given n, $\acute{s}$ is minimum when the rotation angle is equal to half the angle subtended by a side (see right, Fig. 3.13). This angle is the second factor mentioned above, i.e. the worst-case projection angle or worst-case view.

From Fig. 3.13, the size s of each marker face is given by:

$$s_n = D \tan(\pi/n), \tag{3.4}$$

where $n \geq 3$. For the worst-case angle relative to the camera (right, Fig. 3.13), the projection $\acute{s}$ of the largest viewed marker is:

$$\acute{s}_n = s_n \cos(\pi/n) = D \sin(\pi/n). \tag{3.5}$$

This can be visualized in Fig. 3.14 by the plot labeled "s variable." $\acute{s}_n$ has been non-dimensionalized by dividing by body part or bracelet diameter D. It is maximum when n is 3, as long as marker size and bracelet face size is the same and square i.e.,

$$s_n = w_n = h_n = x_n = y_n = D \tan(\pi/n).$$

Note that this is a compound effect of decreasing size and decreasing worst-case projection angle. The effect of changing just the worst-case projection angle while keeping marker size s fixed can also be observed. This is done by setting marker size to some constant number of sides $n = m$. And then the same fixed size s_m is used for all s_n where $3 \leq n \leq m$. In this case, the worst case projected area $\acute{s}$ is given in terms of fixed size s_{fixed} (or s_m) by:

$$\acute{s}_n = s_m \cos(\pi/n) = D \tan(\pi/m) \cos(\pi/n), \tag{3.6}$$

where m is a constant and $3 \leq n \leq m$. This can be visualized in Fig 3.14 by all the curves labeled "s fixed." The "s fixed" curves are illustrated for 4, 6, 8, and 10 sided bracelets (that is $m = 4$, 6, 8, and 10). In each case, it can be seen that the projected size $\acute{s}$ increases as n increases, opposite of the effect when the faces are kept square. In this case, marker size s is usually smaller than the bracelet face width w and the faces are rectangular, i.e.,

$$s_n = x_n = y_n = h_n = h_m,$$

where $w_n > h_n$ when $n < m$, and $w_m = h_m$.

If the bracelet faces are kept square, then minimizing the number of faces is beneficial. If the bracelet height is fixed, then maximizing the number of faces until the faces become squares is beneficial. The tension between these strategies can be quantified in the bulkiness factor. The bulkiness factor is:

$$bulkiness = D_C - D_I, \tag{3.7}$$

$$w_n = h_n = s_n \rightarrow bulkiness \propto \frac{1}{n},$$

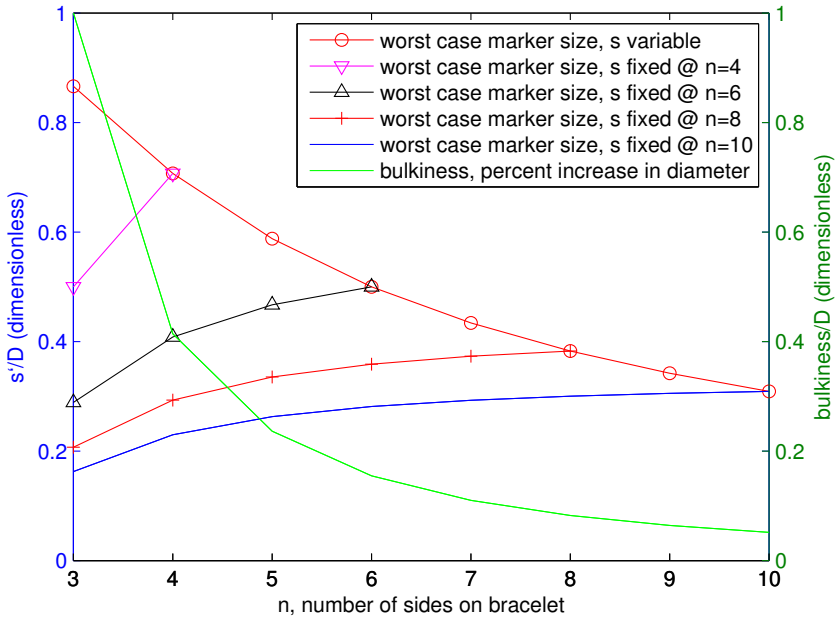


Figure 3.14: "s variable" curve shows the combined effect of decreasing marker size and decreasing worst-case projection angle as n increases (higher $\acute{s}/D$ is better). "s fixed" curves show the effect of decreasing worst-case projection angle only (higher $\acute{s}/D$ is better). The "bulkiness" curve shows the bulkiness factor as n increases (lower $bulkiness/D$ is better). Note that the intersection with the bulkiness factor curve does not signify anything, since it can be scaled to any value.

where D_C is the diameter of the circle circumscribed around the outside of the bracelet, and D_I is the diameter of the circle inscribed inside the bracelet. D_I is the same as D in Fig. 3.13.

The bulkiness factor is illustrated in Fig. 3.14. It has been non-dimensionalized by D so the plot indicates the fraction by which D_C is larger. For example, at n=3, D_C is 100 % larger than D_I, and at n=4, D_C is only about 40 % larger. In terms of the bulkiness factor, ideally it is desirable to maximize the number of sides n, as a larger value of n makes the bracelet less bulky. A larger value of n also has the potential to detect more accurately the centerline of the bracelet (and hence wrist or foot).

Practically, it is desirable to maximize h and y and minimize n to maximize the size of the marker. But it is also desirable to limit the bracelet bulkiness. The intersection with the bulkiness plot in Fig. 3.14 does not signify anything, since the bulkiness factor can be scaled to any value. The largest allowable value of the bulkiness factor depends on feedback from physical therapists on the basis of how much each n sided bracelet interferes with an infant's activities. There is also some concern about potential sharp corners when n is very small (e.g., n=3). Based on discussions with physical therapists, the acceptable bulkiness factor is $bulkiness/D_c \leq 0.236$, i.e., where $n \geq 5$. Physically, this means that the outer diameter of the marker bracelet (D_C) should be at most 23.6 % larger than the diameter of the body part that it is being worn on (D_I or D).

Therefore, to conclude this subsection, the selected bracelet design has five sides. Each side is a square. This is a tradeoff between the maximum visible size at the worst-case viewing angle, and the maximum allowable bulkiness that does not impede an infant's activities.

3.3.5 Dynamic Considerations

For the purpose of capturing infant motion, it is essential that the speed of infant motion does not require shutter exposure times beyond the capabilities of a typical off-the-shelf camera. Typically, camera shutter exposure times can go as low as about 0.05 ms to 0.1 ms and as high as 500 ms [24]. This is important when there is relative motion between a marker and the camera, and does not apply when there is no relative motion between the two.

Camera CCDs are exposed to incoming light for a finite amount of time to capture an image. The lower limit of this time is dictated by the limits of the

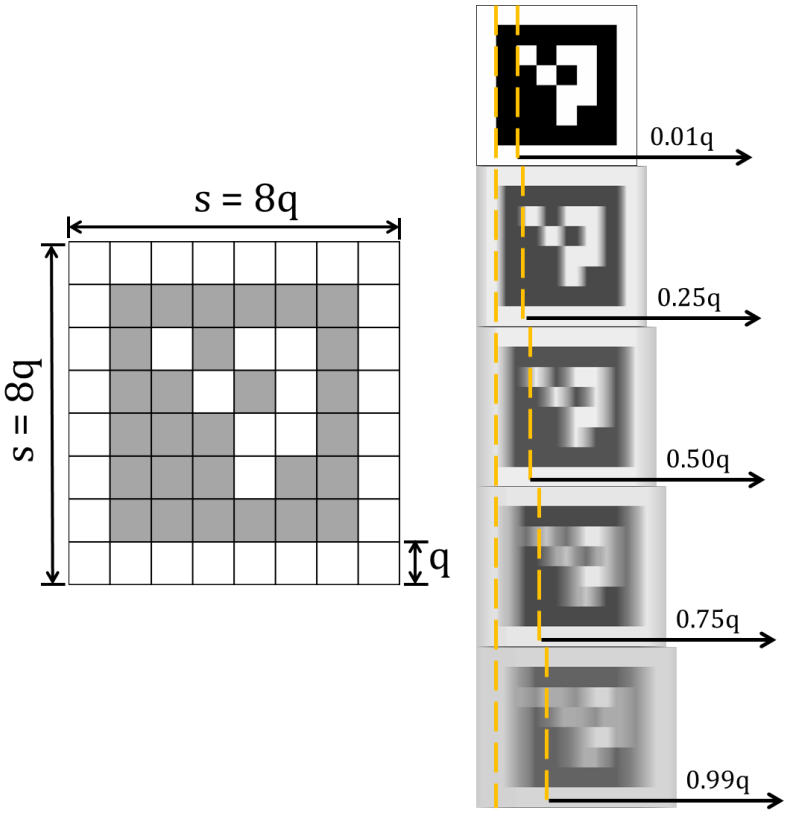


Figure 3.15: Dynamic considerations. Left, MoViT marker and grid dimensions. Right, blurring as the marker moves distance $\gamma q = \gamma \frac{s}{8}$ while the shutter is exposed for a time duration t_{exp}. From (3.8), required shutter exposure time t_{exp} should be within the limitations of the camera electronics. Top right, negligible blurring. Bottom right, extreme blurring.

camera electronics. The upper limit is dictated by the amount of acceptable blurring in the image. At one extreme, the exposure time should be long so that a sufficient number of photons can enter the camera to register a discernible, well-lit image. The fewer the number of photons, the darker the image. If there are very few photons, then no contrast will be registered in the image. At the other extreme, the exposure times should be short so that moving objects have sharp outlines. When the sharp outlines of the marker patterns become too blurred, the patterns cannot be reliably detected by the existing marker detection algorithms in OpenCV.

The limiting factor is the minimum exposure time allowed by the camera electronics. The issue of sufficient lighting can be dealt with by shining more light at the scene. So the shutter exposure time can be decreased as long as sufficient lighting is added to the scene. The faster an object moves, the shorter the exposure time needs to be in order to mitigate blurring effects. At some maximum speed, the minimum shutter exposure time limit is reached.

Assuming a non-zero speed, a marker[11] of size $x = y = s = 8q$ (see left, Fig. 3.15, and also Fig. 3.12, Fig 3.13) would lose its contrasting black-white edges if it travels a distance $\frac{s}{8} = q$ while the camera shutter is exposed. In other words, step changes in the color of adjoining pixels would be negligible. These step changes in color are required by edge detection algorithms in OpenCV for reliable pattern detection. If the camera shutter exposure time is t_{exp} and marker speed is v_m, then, for a marker in motion:

$$t_{exp} \leq \frac{\gamma s}{8 v_m}, \tag{3.8}$$

where $v_m > 0$, $\gamma > 0$, and $s = 8q$. γ is a factor that indicates how far the marker has traveled relative to the size of one grid element $\frac{s}{8} = q$ (left, Fig. 3.15). γ represents the maximum acceptable blurring effect for the marker pattern detection algorithm. If $\gamma = 0.01$ then the marker has traveled $0.01\frac{s}{8} = 0.01q$ and blurring is negligible (top right, Fig. 3.15). If $\gamma = 0.50$ then the marker has traveled $0.50\frac{s}{8} = 0.50q$ and blurring is moderate (center right, Fig. 3.15). If $\gamma = 0.99$ then the marker has traveled $0.99\frac{s}{8} = 0.99q$ and blurring is extreme (bottom right, Fig. 3.15).

This blurring effect in terms of γ is illustrated in Fig. 3.15. Note that 3.8

[11]The MoViT marker pattern is made up of an 8×8 grid of black and white squares.

holds true only for a marker in motion. It is irrelevant and is not defined for a static marker, for which there is no motion blur. γ is limited to a maximum value of 1 in the current implementation where gradient-based edge detection algorithms are used to detect marker patterns. For practical purposes, one reasonable approximation is $0.5 \leq \gamma \leq 1$, depending on the pattern detection algorithm used and camera white balance settings.

3.4 Design Summary

A Monocular Vision-Based Tracking (MoViT) system has been developed. It is a motion capture system for tracking crawling motions of infants and is based on planar pattern markers. The system consists of planar pattern AR markers placed on the wrists and ankles to track them in 3D (see Fig. 3.10). A single camera can be used to track markers in 3D. More markers can be added as desired. The MoViT marker pattern has been derived from the ArUco [34] marker pattern. The MoViT marker contains a 4×4 array of binary squares (each of size q) that define a unique identification code. The marker then has a black border of thickness q on all sides. This surrounded by a white border of thickness q. The entire marker is sized $x = y = 8q$ (see Fig. 3.12). This makes up a single face of a five-sided bracelet. Marker bracelets can be worn at the hands and feet, and a single reference marker can be placed at the back.

Chapter 4

Experiments

In the Chapter 3, a Monocular Vision-Based Tracking system called MoViT was developed in the context of infant crawling motions. The system uses planar pattern markers to track motion. A single camera may be used to track all the markers. Not all markers may be simultaneously visible to one camera, so more cameras may be required. The design requirements for MoViT were identified in Chapter 2.

The next step is to evaluate the MoViT system through physical tests. This chapter broadly covers 4 different physical tests to evaluate performance. In Section 4.1, the utility of the border modification presented in Section 3.3.3 is examined. The objective of this test is to justify this modification. In Section 4.2, the error in position tracking is measured. The objective of this test is to evaluate tracking accuracy for different conditions, e.g., marker size, marker distance, and camera point of view. In Section 4.3 the dynamic limit is examined. The objective of this test is to confirm that the worst-case (longest) shutter exposure time matches the predicted value. Another objective is to confirm that lowering the shutter exposure time helps with marker detection, and

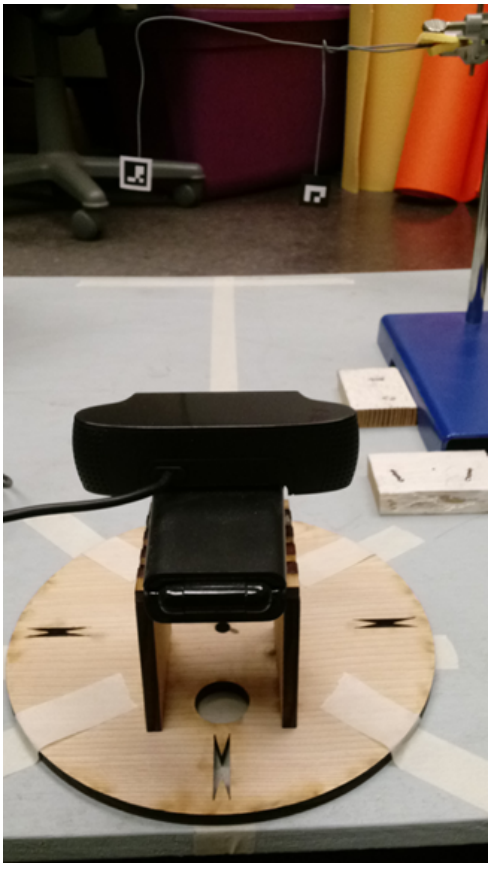
that these exposure times are above the typical minimum exposure time limits for cameras. Finally, in Section 4.4, tracking accuracy is observed for tracking in the frame of reference of a reference marker. The objective of this test is to guide future work and the final implementation.

4.1 Contrasting White Marker Border

4.1.1 Objective

The objective of this experiment is to verify that the modification of the marker by the addition of a contrasting white border is indeed useful. This is done by attempting to detect the original marker design and the new design in front of backgrounds of different colors. The details of white border modification to create the MoViT marker are outlined in Section 3.3.3. The justification for the addition of the white marker border was that the original ArUco planar pattern marker with its black border may not always be distinguishable from different colors in the background, on the clothing, or even adjacent markers on the bracelet (see Section 3.3.2 for bracelet concept). The original design of the marker, like other planar pattern AR markers, assumes that the marker will be printed on a contrasting background (usually white printer paper), and that adjacent markers will be adequately spaced. This ensures that all 4 edges of the marker are detected. Otherwise, even a single missing edge can render the marker undetectable in an image.

(a) Representative camera view



(b) Overview of setup

Figure 4.1: Modified marker border experiment setup illustrating marker placement relative to the camera. MoViT marker on the left and ArUco marker on the right.

Figure 4.2: Modified marker experiment. Panoramic views of three different backgrounds used.

4.1.2 Experimental Setup

The test setup comprised a wooden base with two different markers mounted to a stand. One was an unmodified ArUco marker (marker on the right, Fig. 4.1a), and the other was the modified version with the white border (left, Fig. 4.1a). A USB camera (Logitech, model c920) camera was mounted to the same base at a distance of approximately 500 mm. Fig. 4.1 shows the experimental setup. The camera was used to capture images of the markers. These images were then used to detect the ArUco marker and the MoViT marker. OpenCV 3.1.0 and its companion version of the ArUco library [55] were used for marker detection. The camera resolution was set at 1920×1080 pixels and auto-focus was disabled using Logitech's helper software.

Images were continuously captured and processed as the base was rotated 360 degrees. This process was repeated at three different locations. This changed the background and lighting behind the markers in the camera images. This way, the detection process was tested against backgrounds of a variety of color and lighting conditions. There was no relative motion between the markers and the camera. Fig. 4.2 shows panoramic views of the three different locations.

Table 4.1: Camera calibration parameters obtained for the Logitech C920 camera with 1920×1080 resolution.

parameter	**value (3 sig. figs.)**
f_x (pixels)	1.38×10^3
c_x (pixels)	9.53×10^2
f_y (pixels)	1.38×10^3
c_y (pixels)	5.54×10^2
k_1	1.05×10^{-1}
k_2	-1.76×10^{-1}
p_1	-2.22×10^{-4}
p_2	2.85×10^{-3}

4.1.3 Results

A total of 2800 frames were captured. The detection rate for the ArUco marker was 26.4 %. For the MoViT marker with the white border, the detection rate was 100 %. The lower detection rate for the ArUco marker can be attributed to backgrounds with shadows, and with colors which have a grayscale value close to black. Even if one of the four marker edges blend into the background, such that there is no sharp contrast (when converted to grayscale), then the ArUco marker is undetectable. One example of this blending is Fig. 4.1b.

4.2 Static Accuracy

4.2.1 Objective

The objective of this experiment is to estimate the accuracy in the vicinity of the expected workspace and potential marker positions and orientations. The size of a tracked marker can vary. If a body part is large, then a large marker can be placed on that part. The marker can be located at various distances

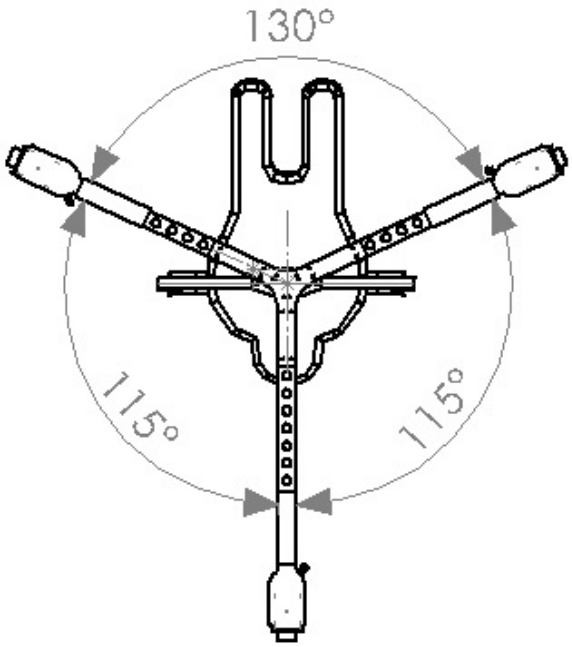


Figure 4.3: Top view of SIPPC-3 robot showing the 'Y' shaped structure with 3 'legs.' The longest 'leg' is about 470 mm long. Any mounted cameras have to be within that structure. Image source: Ghazi et al. [35].

from the camera. It can also be placed at different angles with reference to the camera image plane. This angle is known as the planar angle and is illustrated later in this section (see Fig. 4.5). Finally, a marker can be located at different angles within the camera's field of view. The objective of this experiment is to investigate the tracking error by varying all these conditions.

4.2.2 Experimental Setup

This was a two-part experiment with different conditions, but the same general test setup, which is described below. The test setup comprised a sliding cart mounted on top of a track (see Fig. 4.4). A camera fixture was mounted at one end of the track, and a marker fixture was mounted on the sliding cart, directly in front of the camera fixture (for setting marker distance or radial

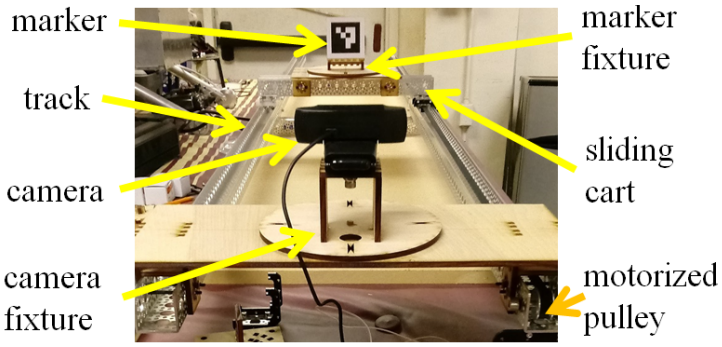


Figure 4.4: Experimental setup to determine system accuracy.

distance). The camera fixture was used to pan the camera right or left (for changing marker angle in camera field of view, see Fig. 4.8). The marker fixture was used to turn the marker right or left (for changing marker planar angle, see Fig. 4.5).

The following guidelines were used to determine the maximum distance from the camera to the marker. If viewed from the top, the SIPPC-3 robot has a 'Y' shaped structure around the infant. This is illustrated in Fig. 4.3. The length of the longest leg of the 'Y' structure is approximately 470 mm. Any mounted cameras have to be within than structure. Therefore, the distance between a mounted camera and a marker of interest on an infant can be up to 470 mm.

For all the tests, the marker or bracelet being tested was moved to within ± 1 mm and the angles were set to within ± 1 degrees of the designated position.

A Logitech USB camera (model c920) was used to capture images. These images were then processed using a laptop to detect the ArUco markers and compute their 3D position. OpenCV 3.1.0 and its companion version of the ArUco library [55] was used. Camera calibration parameters were required for these computations. These were obtained using the OpenCV camera calibration modules and are listed in Table 4.1. The camera resolution was set at 1920×1080

pixels and auto-focus was disabled using Logitech's helper software.

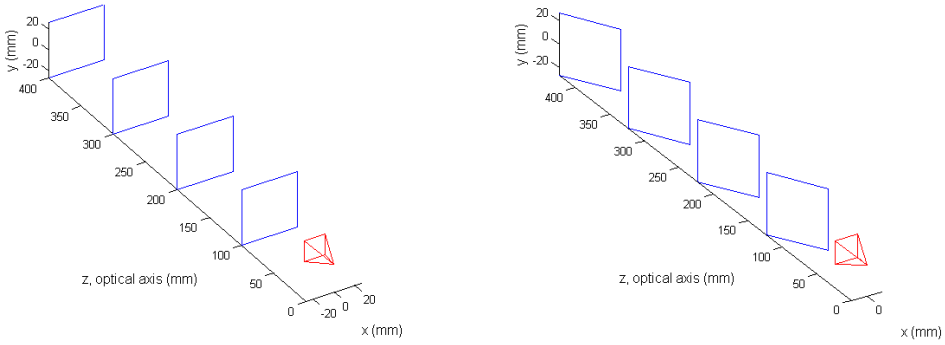
4.2.3 Varying Distance and Planar Angle

In the first experiment, the tracking accuracy for single markers in terms of marker size, planar angle[12], and distance from the camera was investigated. See Fig. 4.5a and Fig. 4.5b for an illustration of the planar angle. The planar angle was changed using the marker fixture shown in Fig. 4.4.

Three different marker sizes were used. The MoViT marker sizes quoted below include the white border (thus the tracked marker was $8q \times 8q$ in size). With the white border included the marker sizes were $x = y = 26.7$ mm, 53.3 mm, and 80.0 mm. This is equivalent to $q = 3.33$ mm, 6.67 mm, and 10.0 mm, respectively. The camera distance was changed from 200 mm to 1000 mm in 100 mm increments. At each distance, the marker was tracked at planar angles of 0, 30, and 60 degrees respectively. For each angle, 20 measurements were taken. After each measurement, the marker was moved away to a different position and then brought back to the desired position. This was done to mitigate potential systematic errors in positioning. For each condition (marker size, distance, planar angle), the mean, lower quartile, and upper quartile for tracked position error were computed.

The minimal camera focus was estimated to be at a point between 100 mm and 200 mm. Images closer than that point were blurred and out of focus. The camera calibration model for computing the pose does not hold true in this region and this introduces error in the marker pose estimation. Therefore, the shortest tracking distance used was 200 mm. Images of objects further than 200 mm were in focus.

[12]The angle between the marker plane and the camera image plane.



(a) 0 degrees (b) 60 degrees

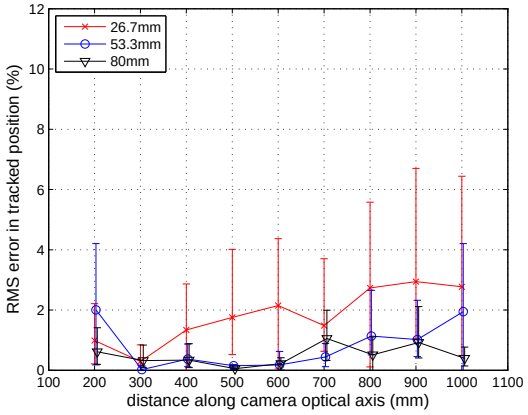
Figure 4.5: Accuracy experiment 1 setup illustrating the marker at different positions and different planar angles (0 and 60 degrees) with reference to the camera image plane.

The results are shown in Fig. 4.6 and Fig. 4.7. Error bars indicate lower and upper quartiles. Note that this experiment, and in fact this chapter, is aimed primarily at investigating 3D position tracking errors. Errors in 3D orientation tracking are not explicitly presented or analyzed. Marker orientation was simultaneously computed, however. The orientation tracking results for this experiment are presented in Appendix C.

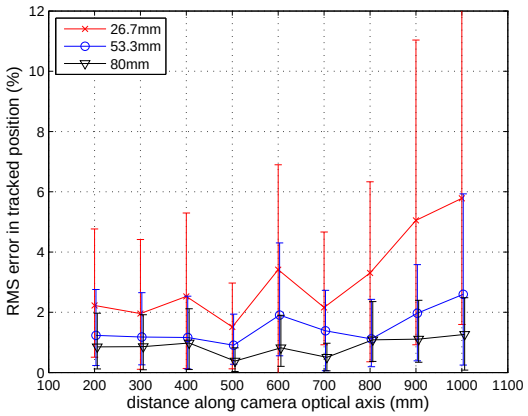
4.2.4 Results: Distance and Planar Angle

For the first experiment for marker accuracy, the results are presented in Fig. 4.6 and Fig. 4.7. The plot for the 26.7 mm marker at 60 degrees does not extend all the way to 1000 mm (Fig. 4.6c, Fig. 4.7c). This is because the size of the marker image decreased to the point that it was undetectable. The plots of percentage RMS error versus distance are shown in Fig. 4.6. Absolute RMS error is shown in Fig. 4.7. The error bars indicate lower and upper quartiles.

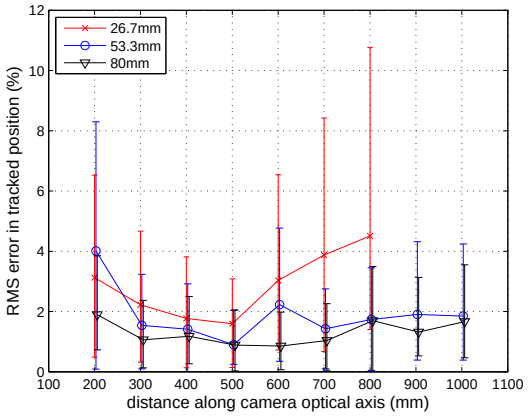
Fig. 4.6 indicates that in general, the tracking error does not seem to be correlated with distance and remains at or below 4 %. The exception is the



(a) 0 degrees

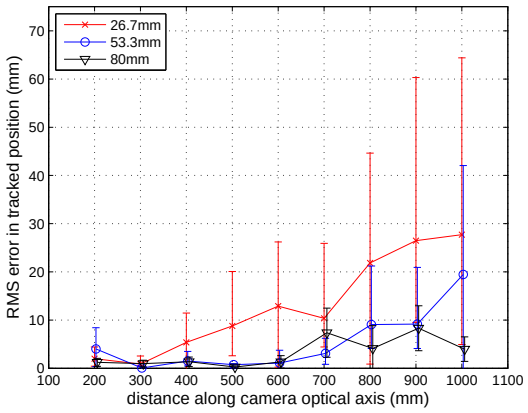


(b) 30 degrees

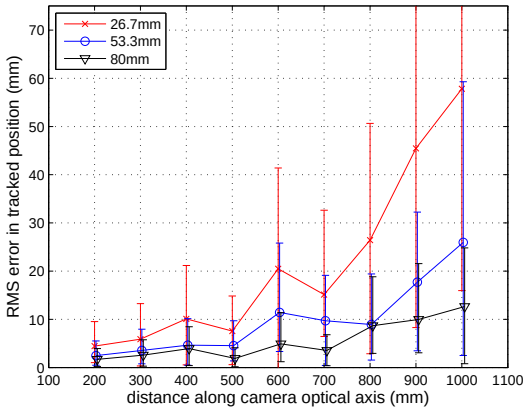


(c) 60 degrees

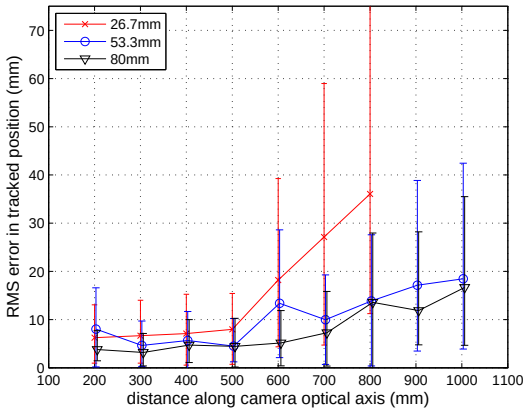
Figure 4.6: Accuracy experiment 1. Percentage RMS position error at planar angles 0, 30, and 60 degrees. Error bars indicate lower and upper quartiles. Marker sizes $x = y = 8q$ were 26.7 mm, 53.3 mm, and 80.0 mm. Sample size $n = 20$.



(a) 0 degrees



(b) 30 degrees



(c) 60 degrees

Figure 4.7: Accuracy experiment 1. Absolute RMS position error at planar angles 0, 30, and 60 degrees. Error bars indicate lower and upper quartiles. Marker sizes $x = y = 8q$ were 26.7 mm, 53.3 mm, and 80.0 mm. Sample size $n = 20$.

increase in error is shown for the smallest marker (26.7 mm) when placed further than 600 mm. This is most likely an anomaly because, at this point, the marker is approaching the limits of detection. It appears that the smaller the marker, the greater the tracking error and hence less accurate the tracked position.

Accuracy does not seem to be significantly affected by planar angle, except for when the the smallest marker (26.7 mm) is beyond 500 mm, approaching the limits of detection. The individual absolute position errors can be seen in Fig. 4.7.

Overall, some noise is to be expected in Fig. 4.6 and Fig. 4.7 due to the discretization of the marker into pixels and due to the nature of edge detection algorithms. For detection of sharply contrasting marker edges, automatically varying thresholds are used. This allows for edge detection under changing lighting conditions, but the exact pixel location may be slightly different each time the camera color balance changes, which can happen every frame, even under constant lighting conditions.

There seems to be an anomalous increase in error at a camera distance of 600 mm for all three marker sizes at a planar angle of 30 degrees (see Fig. 4.6b, Fig. 4.7b). The combination of 600 mm distance and 30 degrees planar angle seems to have a significant error regardless of marker size. An overview of the raw pixel data indicates that data points with larger position errors had different gradients for the top and bottom edges of the markers. Most likely, the perspective projection of a marker edge is more sensitive to variations in edge detection when the marker is at 600 mm and 30 degrees. Alternatively, the pose estimation algorithm may be more sensitive to the marker edge gradient at this particular position and orientation.

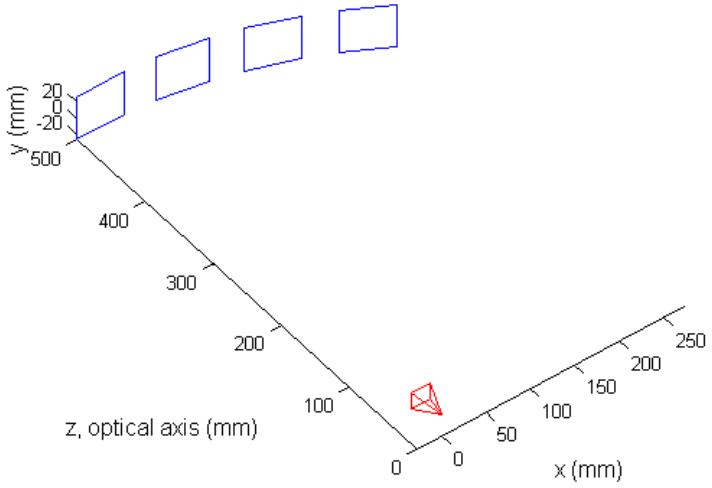


Figure 4.8: Accuracy experiment 2 setup illustrating the marker moving towards the edge of the camera's field of view, away from the optical axis, but at a fixed radial distance from the camera.

4.2.5 Varying Camera View Angle

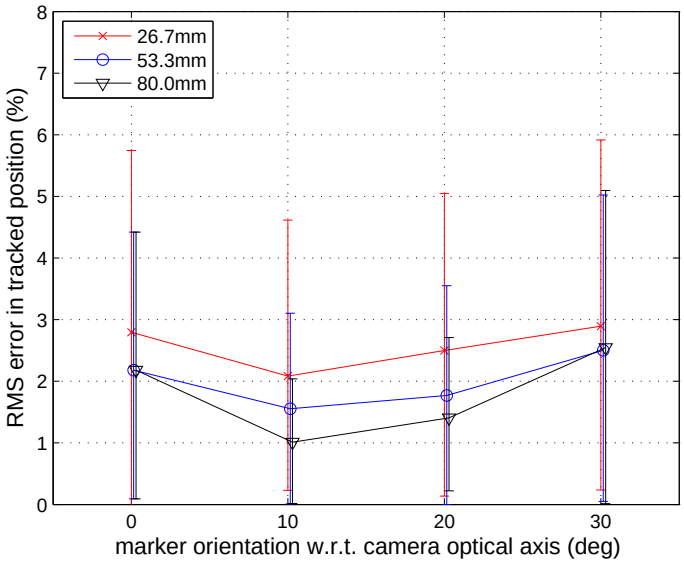
In the second experiment, the change in tracking accuracy was investigated under the following conditions: A marker was moved towards the edge of the camera's field of view, away from the optical axis, but at a fixed radial distance from the camera. The setup is illustrated in Fig. 4.8. The camera was turned from 0 to 30 degrees in 10 degree increments. The angle was changed using the camera fixture shown in Fig. 4.4. These tests were done at a distance of 500 mm for marker sizes $x = y = 26.7$ mm, 53.3 mm, and 80.0 mm ($8q \times 8q$). Each marker size, angle and distance combination was repeated 20 times. The mean, upper quartile, and lower quartile for tracked position error were computed. The results are presented in Fig. 4.9. Error bars indicate lower and upper quartiles

4.2.6 Results: Camera View Angle

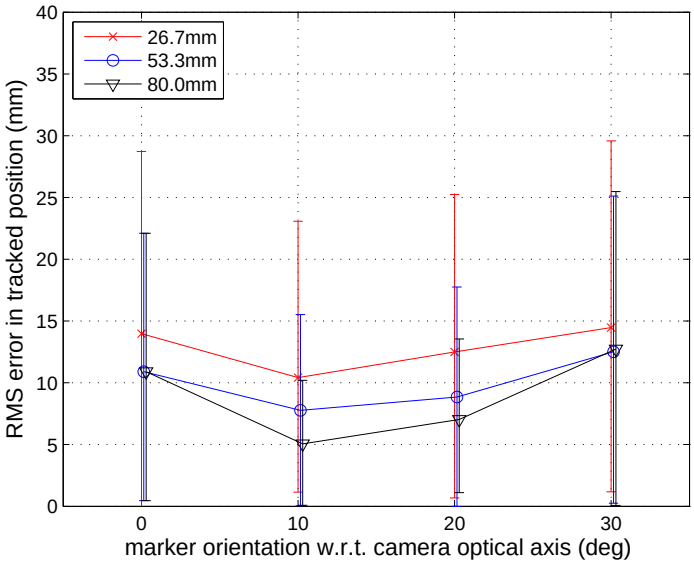
From the results in Fig. 4.9, it can be seen that the further away a marker is rotated, the greater the tracking error. The best accuracy is achieved at 10 degrees rather than the optical axis. This is not a trend but merely an outcome of noise and discretization. Noise is expected due to errors in pixel discretization and changes in automatic color balance. Errors in pixel discretization arise because the boundary between two colors in a scene may fall at the center of a pixel. But because a pixel can only represent one color, the boundary in the image will be shifted by up to half a pixel. Another observation that supports the claim that the curves in Fig. 4.9 do not represent a real trend, is that the absolute error for all three markers at 0 degrees is slightly greater than the error at 500 mm in Fig. 4.7a, which is exactly the same test condition (marker 500 mm along optical axis, marker plane parallel to camera image plane). Moreover, from Fig 4.9a, the overall change in percentage error is very small, i.e. from 1 % to 3 %.

Further evidence to support the claim that the apparent trend in Fig. 4.9 is not significant is provided in Appendix B. Appendix B contains supplemental results. The results are based on this experiment, repeated for two more cameras (Logitech C920, Logitech C615). The results from both cameras show no discernible trend.

To conclude, there is no correlation between tracking error and marker orientation with reference to the camera optical axis.



(a) percentage error



(b) absolute error

Figure 4.9: Accuracy experiment 2 results (camera view angle). Error variation as the marker is rotated away from the optical axis of the camera. Error bars indicate lower and upper quartiles. The radial distance to the marker is 500 mm. Sample size $n = 20$.

4.3 Dynamic Limitations

4.3.1 Objective

The main objective of this test is to confirm that the worst-case (longest) shutter exposure time matches the time predicted by (3.8). Another objective is to confirm that lowering the shutter exposure time helps with marker detection, and that these exposure times are above the typical minimum exposure time limits for cameras. In Section 3.3.5, the dynamic limitations were discussed. For a given marker speed, there is an upper limit to the camera shutter exposure time t_{exp}. At this upper limit, the marker becomes blurred beyond recognition by OpenCV (where ArUco pattern detection is based on detecting contrasting edges). For camera shutter exposure times below this limit, the blurring effect decreases and the chances of detection increase. The hard limit for the shorter exposure time is the camera hardware and this exposure time must be greater than the limit imposed by the hardware. All these effects limit the maximum speed of a wearable marker that can be detected. Given a marker speed v_m, the camera shutter exposure time t_{exp} can determined by (3.8).

4.3.2 Experimental Setup

A marker of size 26.7 mm was moved at 0.5 m/s. The marker was mounted to a motorized wheel of radius 900 mm. The wheel was rotated such that the tangential speed of the marker was 0.5 m/s. The motion of the marker in this case is an approximation of linear motion, with the outer edge of the marker pattern traveling slightly faster than the center (about 1.5 %).

From (3.8), the maximum shutter exposure time is $\frac{\gamma \times 26.7\ mm}{8 \times 500\ mm/s} = 6.67$ ms

(maximum when $\gamma = 1$). Images of the moving marker were captured for shutter exposure times 3.60 ms through 6.60 ms at 1.00 ms intervals. These images were processed using the ArUco AR module in OpenCV to check if the marker could be detected by the algorithm [34].

The Logitech C920 webcam, which was used for the other experiments, was not used for this particular experiment. The LGS configuration utility for the Logitech C920 webcam does not have firmware support to control shutter exposure time. This is fairly typical of modern webcams. Therefore, shutter exposure tests were carried out using a Raspberry Pi 3 with the Arducam OV5647 camera at the same resolution as the other experiments (1920×1080 pixels). This setup allows manual control of shutter exposure time.

4.3.3 Results

As expected, the detection rates dropped sharply in the vicinity of $t_{exp} = 6.67$ ms. At at $t_{exp} = 5.6$ ms, the detection rate was 0 %. By the time $t_{exp} = 3.6$ ms, the detection rate was 73 %. Representative images of the blurred markers are shown in Fig. 4.10. This agrees with the predicted maximum shutter exposure time of 6.67 ms. At this exposure time, the detection rate should theoretically be zero. For shorter exposure times, the detection rate should increase. Note that absolute minimum camera shutter exposure times are typically about 0.05 ms to 0.1 ms [24]. There is some margin for the shutter exposure time to go below the 3.6 ms used for the experiment.

An interesting observation here was that a shutter exposure time of 1.00 ms was also attempted to get a benchmark result. All images of the marker were very sharp, i.e., with no apparent blurring. But at this point, the lighting of the

Figure 4.10: Experiment to verify predicted maximum camera shutter exposure time for a marker moving at 0.5 m/s. Left, representative marker image when it is detectable. Right, representative marker image when it is not detectable.

experiment setup was not enough to light up the image very well. The contrast was not very desirable. Although the marker was visible to the human eye, it was not detectable by the ArUco software 100 % of the time. The detection rate was closer to 75 %. This is not a limitation of the MoViT mocap system. This is a limitation that can easily be overcome by using better lighting to light up the scene, or by using a more sophisticated detection algorithm. Better lighting may even improve results for the longer shutter exposure times.

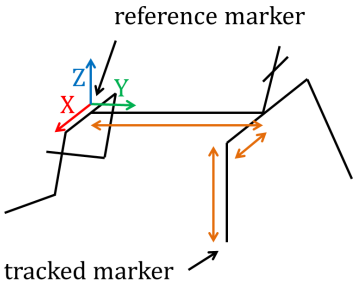
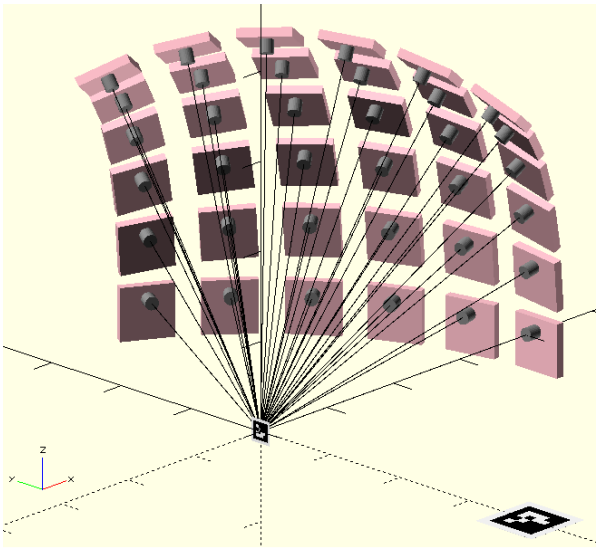


Figure 4.11: Expected location of reference and target markers. Kinematic skeleton of an infant in the prone position. On a 9-11 month old infant, the target marker would be at (105.5, 274, -252). Based on sizing information derived in Table 2.1.

4.4 Accuracy with Reference Marker

4.4.1 Objective

The objective of this reference experiment is to gather preliminary data on tracking performance in more realistic conditions. These data will also help guide the implementation and direction of future work. Practically, a marker will be tracked in the reference frame of a reference marker placed on the lower back of the infant. This means that the total tracking error comprises position errors from two markers as well as any orientation (angular) errors from the reference marker. Also, there is the question of whether the relative orientation of the two markers affects the error. This will help with the placement of the reference marker for the final implementation. Finally, there is the question of performance of the tracking process when the camera is placed at different view points.



(a) camera sweep pattern

(b) same planes

(c) perpendicular planes

Figure 4.12: Setup for reference marker experiment. A target marker was continuously tracked in the frame of a reference marker as the camera was moved in a sweeping pattern from the right to the front and from marker level upwards. (a) Approximate camera sweep pattern. (b) Marker planes parallel. (c) Marker planes perpendicular.

4.4.2 Experimental Setup

A reference marker (size 80 mm) and a target marker (size 26.7 mm) were placed along a 600 mm graduated level. The target was placed at coordinates (0, 387, 0) mm (see reference frame in Fig. 4.11). This simulates the distance between a reference marker placed on the lower back and a target marker on the wrist. On a 9-11 month old infant, the target marker would be located at (105.5, 274, -252) mm (see Fig. 4.11). The Euclidean distance between the reference and target marker placed this way is 386.9 mm. These data are based on infant sizing information derived previously in Table 2.1.

In one test, both markers were in placed parallel to each other, i.e., on the same plane (Fig. 4.12b). This simulates a reference marker placed vertically on an infant's back while a target marker is on the wrist. In a second test, both markers were in perpendicular planes (see Fig. 4.12a, Fig. 4.12c). This simulates a reference marker flat on an infant's back while a target marker is on the wrist. Target marker position was continuously tracked in the reference marker frame as the camera was swept around while keeping both markers in view. The sweep pattern is illustrated in Fig. 4.12a.

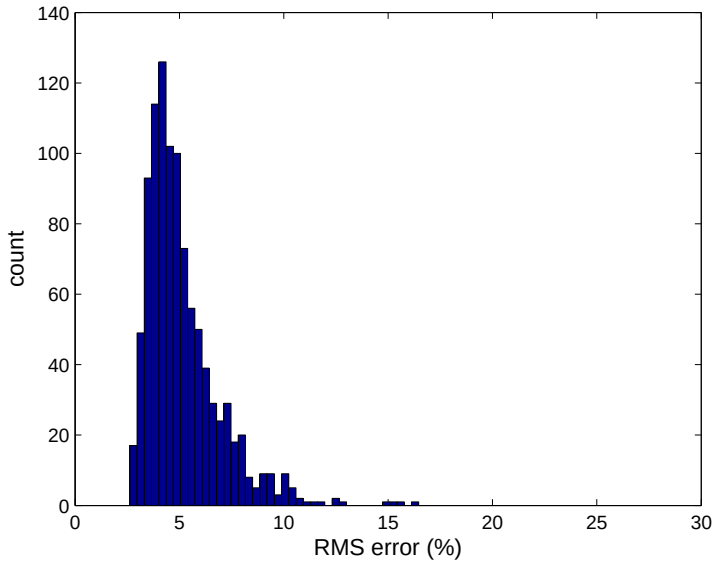
For the starting position for the parallel markers, both markers faced the camera (Fig. 4.12b). For the starting position for the markers on perpendicular planes, the target marker faced the camera while the reference marker faced upwards (Fig. 4.12a, Fig. 4.12c). For the camera sweep pattern, the radial distance of the camera and target marker was kept constant at approximately 400 mm. The sweeping pattern was from the right side of the system to the front of the system, a rotation of about 60 degrees. Subsequent sweeps were made such that the camera was moved higher up and tilted down. The upward

sweep also spanned about 60 degrees. This simulates the camera rotating from the side of the baby towards the front, and moving from marker level up to a level higher than the baby.

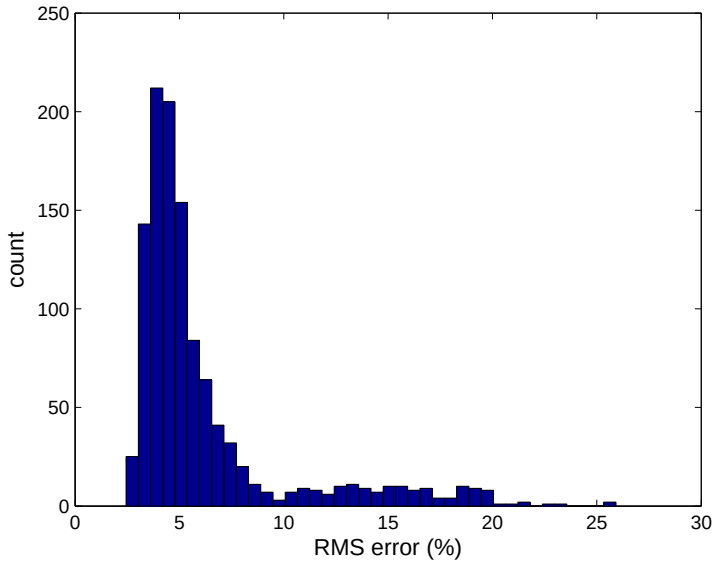
4.4.3 Results

Histograms for percentage RMS position error are shown in Fig 4.13. This is a percentage of the Euclidean distance between the reference and target markers, i.e., 386.9 mm. Fig. 4.14 shows the same results in terms of RMS position error in mm. For both configurations, the mean RMS error is comparable (5.15 % and 6.18 %, or 19.9 mm and 23.9 mm). This provides an insight into the typical error based on the combined position error from the two markers, and angular error from the reference marker. Fig 4.13b and Fig. 4.14b show that the distribution for the perpendicular plane setup is skewed, indicating a tendency to have large tracking errors some of the time. The 95th percentile for the perpendicular setup is 16.1 % (62.2 mm), while for the parallel plane setup, it is 8.66 % (33.5 mm) as a percentage of the distance between the two markers. This indicates that for some tracking positions, the perpendicular plane setup can have significantly poor performance.

The most likely cause for the skewness in the perpendicular plane setup results (4.13b and Fig. 4.14b) is that there is a specific row of camera positions at which the tracking error is significantly larger (see Fig. 4.12a for the rows of positions). If the individual error from each marker is proportional to the planar angle between the marker and the camera image plane, then this would be the row where the camera plane is oriented at 45 degrees with each marker. An alternative explanation is that these larger errors occur at a specific column

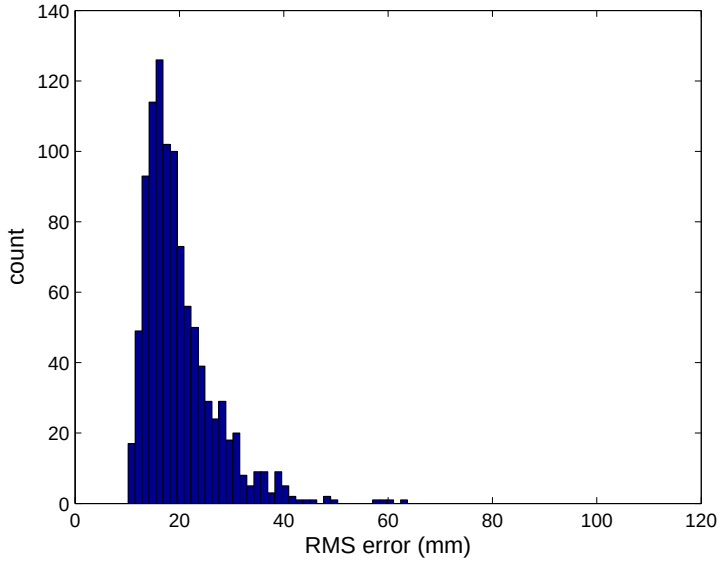


(a) parallel, $n = 999$

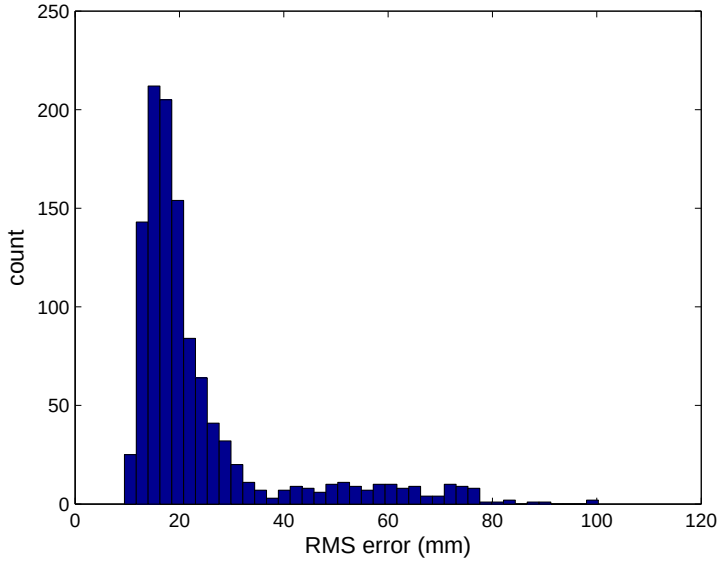


(b) perpendicular, $n = 1148$

Figure 4.13: Results for reference marker experiment (percentage). (a) Planes of markers parallel. Mean 5.15 %, 95th percentile 8.66 %. (b) Planes of markers perpendicular. Mean 6.18 %, 95th percentile 16.1 %.



(a) parallel, $n = 999$



(b) perpendicular, $n = 1148$

Figure 4.14: Results for reference marker experiment (absolute). (a) Planes of markers parallel. Mean 19.9 mm, 95th percentile 33.5 mm. (b) Planes of markers perpendicular. Mean 23.9 mm, 95th percentile 62.2 mm.

of positions in Fig. 4.12a, perhaps one where the camera is equidistant from the two markers. Both the above explanations assume that both markers affect the error equally. It is also possible that the larger errors depend on the planar angle or position of the target marker alone.

Overall, this experiment has provided a useful insight into the expected tracking error. Overall, the tracking error is higher, which is expected. Tracking error can become significantly higher for the perpendicular configuration, i.e., when a reference marker is placed flat on the back. In terms of implementation, this means that the current design of placing the reference marker flat on the back could be problematic. Further experiments are required to identify whether the perpendicular marker placement is problematic for all camera positions or or only a subset of those positions. Other reference marker placement designs should also be explored.

4.5 Conclusion

From Section 4.1 it is clear that the contrasting white marker border for the MoViT marker is an improvement that allows seemingly 100 % detection rates in the presence of other background colors and objects. The results from Section 4.2 show that for the expected marker distance from the camera (approximately up to 500 mm), the 3D position tracking error is less than 4 % in the worst-case. Based on results from Section 4.3, the maximum predicted camera shutter exposure time for tracking a marker moving at 500 mm/s is consistent with the theoretically predicted time of 6.67 ms. As predicted, detection rate increases from zero as the shutter exposure time is reduced (73 % at 3.6 ms). All these are well within the specifications of typical cameras. Finally, in Section 4.4,

a target marker was tracked by using a reference marker. The combined tracking error is approximately 5 % (or 19.9 mm). This is based on a marker-marker distance of 387 mm, and marker-camera distance of approximately 400 mm.

The objective of the experiments in this chapter is to ultimately compare the performance of the MoViT mocap system with the performance requirements determined in Chapter 2. According to these requirements, the smallest possible crawling motion is expected to have a displacement of 74.6 mm. The combined error of 19.9 mm with a reference marker is sufficient to detect this motion. According to the performance requirements, the fastest possible movement of the marker is expected to be around 304 mm/s[13]. Tracking at this speed is within the limitations of typical cameras, though extra lighting might be required. This was demonstrated by detecting a marker moving at 500 mm/s.

[13]This is a 95th percentile value, so the top 5 % of peak speed values may be much higher, and the marker may not necessarily be tracked at all of those top 5 % speeds.

Chapter 5

A Model to Predict Performance of the MoViT System

Chapter 4 presented some preliminary experiments to evaluate the performance of the MoViT system developed in Chapter 3. In this context, the performance of the system is defined as the tracking accuracy and camera shutter exposure limits. This chapter presents a model to predict the performance without having to physically test the system.

The need for a performance prediction model for the MoViT system is outlined in Section 5.1. This is followed by a survey of works that have evaluated the 3D tracking performance of planar pattern AR markers in Section 5.2. The MoViT marker used in the MoViT system is one such planar pattern AR marker (based on the ArUco marker system). Section 5.3 develops the performance model for the MoViT system. In Section 5.4, three use cases are presented to highlight the capabilities of the model. In Section 5.5, the error predicted by one of the use cases is compared with the error measured from the physical testing.

5.1 The Need for a Model

The tracking performance depends on several factors, e.g., the size and position of the planar pattern markers and the properties of the camera. Therefore, the results presented in Chapter 4 apply to the specific test conditions and the type of camera used. Those results are not necessarily informative if there is a requirement to evaluate the performance of a different configuration, for example with different marker sizes or a camera with a different resolution. It is not feasible to experimentally determine the performance of every variation of the system.

Consider the following example to explore the feasibility of physical testing of the system. For the first experiment in Section 4.2, there were 3 different marker sizes. For each marker size, there were 3 different planar angles. For each marker size at a given planar angle, there were 8 different positions. Each position was set up 20 times. This combination gives $3 \times 3 \times 8 \times 20$ or 1440 different size/angle/position setups. If setting up and logging each size/angle/position instance took 20-30 seconds, then the total time for this experiment is 8-12 hours. If for some reason, 3 more marker sizes were to be tested, then that is another 8-12 hours of testing. That is a lot of testing to obtain an estimate of the error.

An even more challenging testing scenario is related to testing for certain motions of an infant. If the system is tested with a live infant, then some sort of benchmark system will be required. Typically, infared point marker vision systems are very accurate. If an infrared point marker system is used for benchmarking, then a considerable amount of time and effort is required for setting it up. An infant's motion may not be exactly as desired and the

experiment may have to be repeated several times over. Crawling infants cannot be asked to make specific types of motions. They will move as and when they feel like it.

Thus, it is more practical to have a model. Such a model could be used to predict the performance of this planar pattern tracking system based on the different variables involved. The motivation is that this should be useful in evaluating and comparing the performance of variations of the MoViT system developed in Chapter 3. More importantly, it could save a significant amount of time in development and evaluation.

5.2 Survey of Evaluations of Planar Pattern Marker Tracking

The following is an overview of work done in evaluating planar pattern marker tracking. The intention is to discover potential models that may have been developed and could be applicable to the MoViT system.

Much work has been done to evaluate planar pattern AR marker systems. Mostly, the emphasis is on performance aspects related to marker detection, e.g., in work done by Zhang et al. [97], Fiala et al. [28], and Agnus et al. [9]. Such aspects include positive or negative detection rate, pixel error in detected features, and processing time taken for detection. Cesar et al. [25] evaluated AR marker limitations underwater, such as smallest detectable marker size, largest possible distance, and largest possible planar angle. La Delfa et al. [23] also evaluated AR marker detection limitations for different sizes and camera distances.

Others have investigated accuracy in planar pattern marker tracking, but much in the manner of Chapter 4 they have focused only on evaluating specific setups. The results published by La Delfa et al. [23] were applicable only to specific cameras. In fact, La Delfa et al. [23] used 3 different cameras and resolutions for evaluating 3 different types of AR markers, so comparison is difficult. López-Céron et al. [44] have done an excellent analysis of position and angle error for an AR marker. They generated a simulation in the Gazebo simulator and validated it experimentally. Unfortunately, these results hold true only for a specific marker size and a specific camera. Olson et al. [53] used a ray tracer to generate simulated images and then used those images to predict tracking accuracy of the position at difference distances and of the angle at different angles. But their model and results applied to a specific camera and marker size. Furthermore, their camera model, a 400×400 pixel pinhole camera model, was not representative of a real-world camera.

From the above review of previous work in this area, it follows that the work done in evaluating planar pattern marker tracking has been been either limited to detection, or for a very specific set of parameters. Where some aspect of modeling has been involved, it has been limited to fixed design parameters. Therefore, there do not seem to be any models directly applicable to the MoViT system.

5.3 Development of Model

This section describes the development and features of the MoViT model. It is not limited to the MoViT system, but it does incorporate some features that go beyond simply tracking a planar pattern marker in 3D. The model estimates

the tracking accuracy and shutter exposure time limit.

Major factors that affect tracking accuracy are projective geometry, camera lens distortion, and pixel noise in the marker image. The shutter exposure time limit is dictated by the blurring of the image of the marker while it is in motion relative to the camera. The longer the exposure time, the greater the blurring of the image. For a given marker speed, there is an upper limit to the exposure time. This limiting shutter exposure time interval is one where the marker pattern has been blurred beyond recognition (see Fig. 3.12 for marker pattern). At this point, a marker detection algorithm will be unable to detect the defining edges of a marker pattern in the camera image (see right, Fig. 4.10). More details are provided in Section 5.3.3.

This model can be used for: 1) a marker in a predefined workspace; and 2) for a marker following a limb trajectory in space. The required parameters include camera intrinsic parameters, distortion parameters, marker size, and a set of marker poses in the workspace. In the case of the limb trajectory option, joint angle trajectories and the placement of the marker on the surface of the limb are required. In this case, the marker workspace need not be explicitly defined because it is computed using the limb kinematics. An outline of how the model works is as follows:

1. Compute marker pose over a trajectory or space
2. Compute projected camera image
3. Estimate the marker pose from the image
4. Compute the pose error
5. Compute camera shutter exposure limit

After computing the projected camera image (pixel coordinates of the four marker corners), the camera parameters are used to solve the perspective n-point (PNP) problem to estimate the marker pose in 3D. The model uses the Open Source Computer Vision Library (OpenCV [54]) to solve the PNP problem. The PNP problem is a classic computer vision problem and there are several approaches to solving it. The approach used for this model is the default approach implemented in OpenCV called "solve PNP iterative" [56]. According to OpenCV documentation, this approach finds the pose that minimizes the reprojection error in the pixels. The reprojection error in pixels is defined as the sum of squared distances between the pixel coordinates of the marker and the pixel coordinates projected using the estimated pose. The iteration method used is based on Levenberg-Marquardt optimization [43].

There are 3 main components to the model: 1) the marker model, 2) the kinematic model, and 3) the camera model. The marker model defines the geometry of the marker in terms of 4 corners. The kinematic model transforms the geometry of the marker as it moves with a moving limb. The camera which projects the 3D coordinates of the marker onto the 2D camera image plane. Details of these three component models are provided in the following subsections. Limb trajectories are computed at 100 Hz. Based on some infant crawling data from Righetti et al. [72], this is sufficient to model the kinematics. For comparison, Southerland's inertial suit system captures motion data at 50 Hz.

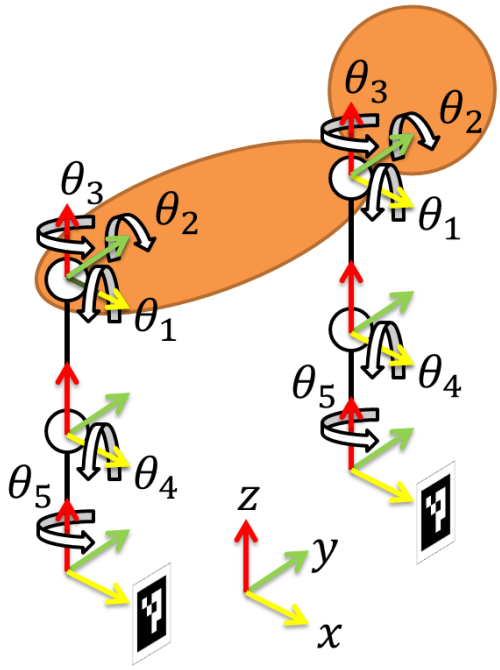


Figure 5.1: The limb kinematics model along with the marker model in the neutral pose. A head and torso are shown in the background for illustration. The convention is the same for either limb (arm or leg).

5.3.1 Marker Model

The marker model transforms marker coordinates from the marker body frame to the end effector frame. Physically, this is the equivalent to placing the marker on a bracelet around the wrist or ankle. This involves a translation (T_3) and a rotation (θ_5) about the limb axis. Marker size needs to be specified. See Fig. 5.1 for an illustration. The transforms are listed in Table 5.1.

It must be emphasized that this model is not limited to the 8×8 grid MoVoiT marker (Fig. 3.12). Any square-shaped marker may be represented, e.g., ARToolKit [41], AprilTag [53], and ARTag [27]. A rectangular-shaped marker may be used with minor modification. The MoViT system requires a border but the marker code/pattern and recognition algorithm can vary.

Table 5.1: Transforms for the kinematic model, listed in the order they are applied in the kinematic chain (arm or leg).

transform	**description**
rotation R_1	θ_1, about x, flexion/extension
rotation R_2	θ_2, about y, abduction/adduction
rotation R_3	θ_3, about z, lateral/medial rotation
translation T_1	along upper arm or thigh
rotation R_4	θ_4, about x, flexion-extension
translation T_2	along forearm or lower leg
rotation R_5	θ_5, about z, marker placement
translation T_3	along x and z, marker placement

5.3.2 Kinematic Model

The kinematic model transforms the marker coordinates from the end effector frame to the root frame of the limb. Physically, it represents marker pose as a result of the limb joint angles. The limb is approximated by a two-link, 4 degrees-of-freedom (DOF) kinematic chain. The first joint is a 3 DOF ball joint and the second joint is a 1 DOF hinge joint. The ball joint is represented by a series of three 1 DOF hinge joints. Fig. 5.1 shows the neutral pose of the kinematics model in the context of a crawling body. The transforms are listed in order in Table 5.1. The neutral or reference pose of a limb is defined with both kinematic links (limb segments) pointing downwards (see Fig. 5.1). The x-axis is to the right, the y-axis is to the front, and z-axis is upwards. Rotations about the x, y, and z axes represent flexion/extension, abduction/adduction, and lateral/medial rotation respectively.

For a point in the end-effector frame $P^{END} = [x\ y\ z]^T$ the transform from the end effector frame to the root frame is given by:

Table 5.2: Infant sizes used in the kinematic model. These data are a subset of the data in Table 2.1 which were derived from anthropometric data from Snyder et al. [79].

	3-5 months (mm)	**6-8 months (mm)**	**9-11 months (mm)**
upper arm	123	131	145
forearm	92	100	107
thigh	101	111	124
lower leg	103	117	134
wrist diameter	32	33	34
ankle diameter	37	39	41

$$P^{ROOT} = R_1 R_2 R_3 (T_1 + R_4 (T_2 + P^{END})) \tag{5.1}$$

where R_i represents a rotation and T_i represents a translation. The definitions of the transforms are listed in Table 5.1. If the transforms in the marker model are included, then the full transform from the marker frame to the root frame becomes:

$$P^{ROOT} = R_1 R_2 R_3 (T_1 + R_4 (T_2 + R_5 (T_3 + P^{MKR}))) \tag{5.2}$$

Since the primary motivation for the MoViT system is capturing infant motion, infant body sizing information has been predefined in the model (see Table 5.2). These data are a subset of infant sizing data in Table 2.1 which were derived from anthropometric data in Snyder et al. [79]. This simplifies usage since only joint angle data and the age group is required to simulate motion. This is in no way a limitation of the model. Sizes for other groups of individuals can also be used.

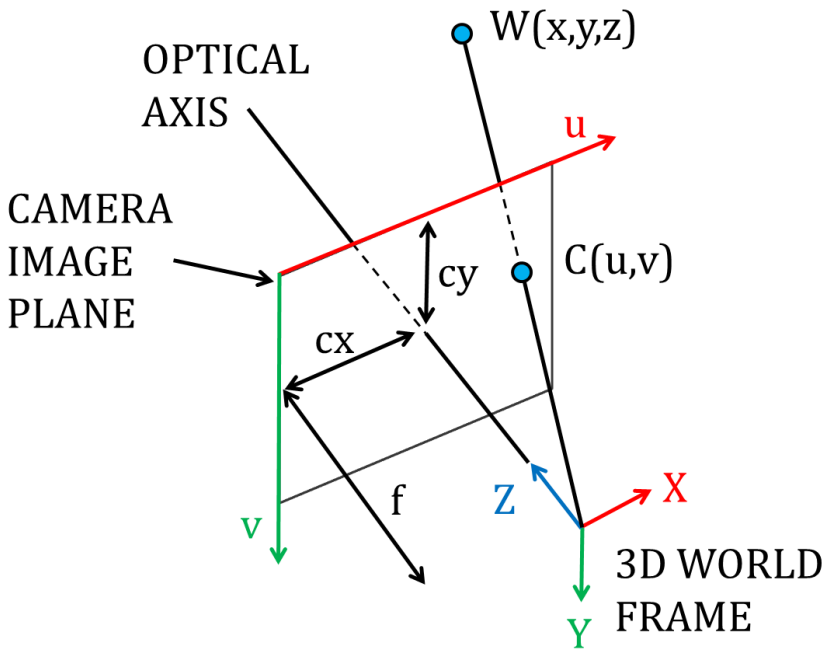


Figure 5.2: Illustration of the pinhole camera model.

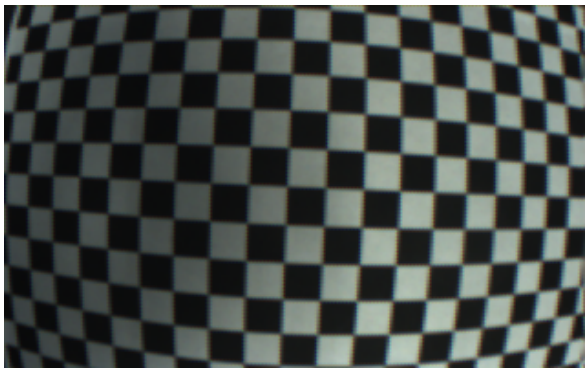
If the marker is to be used in the context of a predefined workspace rather than on a moving limb, then the limb and marker model transforms, defined in Table 5.1, are replaced by a 3D rotation and translation:

$$P^{ROOT} = R(T + P^{MKR}) \tag{5.3}$$

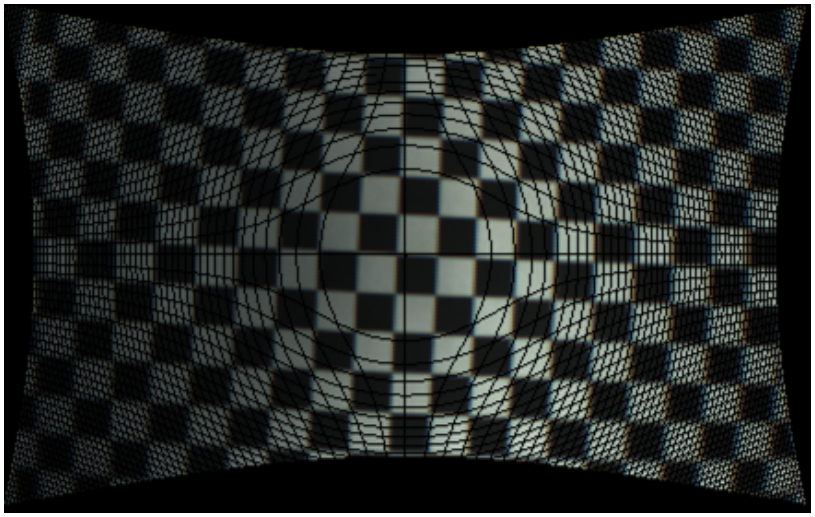
5.3.3 Camera Model

The camera model projects a real world point $W = [x\ y\ z]^T$ (in mm) to the camera image plane $C = [u\ v]^T$ (in pixels) (see Fig. 5.2). It is based on the pinhole camera model modified for lens distortion effects. Given:

$$x' = x/z, \quad y' = y/z \tag{5.4}$$



(a) with radial distortion



(b) corrected for radial distortion

Figure 5.3: Example of positive radial distortion. Based on the image of a regular checkerboard pattern. By using distortion models, an ideal image can be distorted to get a camera image, or a camera image can be undistorted to obtain the ideal image. (a) Image with distortion effects. (b) Image after applying correction using a radial distortion model.

the radial and tangential distortions can be approximated by:

$$x'' = x'(1 + k_1 r^2 + k_2 r^4) + 2p_1 x' y' + p_2(r^2 + 2x'^2) \tag{5.5}$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2y'^2) + 2p_2 x' y' \tag{5.6}$$

where $r^2 = x'^2 + y'^2$. k_1 and k_2 are radial distortion coefficients, and p_1 and p_2 are tangential distortion coefficients. More details on lens distortion can be found in the OpenCV documentation for camera calibration [56]. Finally, the pixel coordinates $C = [u\ v]^T$ can be computed by:

$$u = f_x x'' + c_x + \delta_x \tag{5.7}$$

$$v = f_y y'' + c_y + \delta_y \tag{5.8}$$

where f_x and f_y are the focal lengths, c_x and c_y are the principal points, and δ_x and δ_y are Gaussian noise. δ_x and δ_y account for the fact that the edge detection algorithms detect the marker outline slightly differently in each frame. Some of the factors that can influence this are difference in lighting, automatic white balance, and automatic color balance.

The second part of the camera model computes the limiting shutter exposure time. There are two assumptions: 1) rolling shutter effects are negligible, and, 2) the marker motion is largely parallel to the camera image plane. The reasons for assuming parallel motion are explained in Section 5.3.4.

As discussed in Section 3.3.5 the shutter exposure time is given by (3.8). The shutter exposure time is limited by the speed of the marker (relative to

the camera) and the physical size of a single grid element q on the grid pattern (see Fig. 3.12). The marker grid pattern will be blurred beyond recognition if it travels a distance q, relative to the camera, while the shutter is open. For the marker in Fig. 3.12, $q = x/8 = y/8$ because it is an 8×8 grid. But a different marker may use a different grid size.

5.3.4 Shutter Exposure Time: Assumptions

The following is an explanation of why the marker motion is assumed to be largely parallel to the camera image plane. For comparable blurring effects, motion normal to the camera image plane would have to be much faster than motion parallel to the camera image plane. In other words, blurring effects due to motion parallel to the camera image plane are observed first, at much lower speeds.

Consider the case of motion normal to the camera image plane, with the direction of motion being towards the camera. For a marker pattern to be completely blurred when moving normal to the camera image plane, a grid element $q \times q$ would have to increase to size $2q \times 2q$ (grid element shown in Fig. 3.12). That is to say, the marker would move so fast that a single grid element would appear to grow to the size of 4 grid elements. The apparent marker size would have to be doubled during the shutter exposure time t_s. For the apparent size to be doubled, the distance between the marker and the camera would have to be halved. This means that the marker would have to travel a distance that is equivalent to half the distance between the marker and the camera. For the camera used in the last chapter, the shortest usable target distance from the camera was about 200 mm. In this case, the marker

would have to move a distance of about 100 mm during the shutter exposure time t_s. This is analogous to the minimum expected normal speed that could cause excessive blurring. Note that this distance or speed would be larger if the marker is located further away. This distance or speed will also be larger if the marker is moving away from the camera, since in that case the distance would have to be doubled.

Now consider the case of motion parallel to the camera image plane. A marker would have to move a distance q to become completely blurred. The largest marker size used in the last chapter was 80 $\times$ 80 mm (equivalent to $8q \times 8q$). This marker would have to move 10 mm parallel to the camera image plane during the shutter exposure time t_s in order to become completely blurred. This is analogous to the maximum expected planar speed that could cause excessive blurring. For smaller markers this planar speed (parallel to the camera image plane) would be lower.

Blurring effects due to rotation of a marker are also less problematic. This is because in order to cause a significant amount of blurring, rotation speed must be tens of radians per second. This is much higher than typical movements expected of infants (less than 5 rad/s, according to data from Smith et al. [78]).

This proves that the blurring effects from motion parallel to the camera image plane are dominant. From both the above scenarios, the speeds are as follows. In the case of blurring effects due to normal motion, the speed at which this happens is at least $\frac{100}{t_s}$ mm/s. In the case of blurring effects due to planar motion, the speed at which this happens is at most $\frac{10}{t_s}$ mm/s. In other words, blurring effects due to planar motion (parallel to the camera image plane) become apparent at much lower speeds, as compared to blurring effects due to normal motion (normal to the camera image plane). Therefore, the limiting

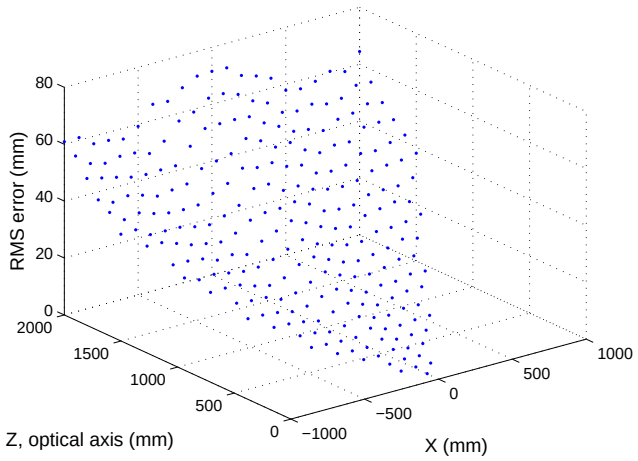


Figure 5.4: Using our model to predict tracking accuracy in a specific workspace. Position tracking accuracy was estimated for different marker positions on the horizontal plane. Camera at (0,0,0).

factor is the planar speed, and blurring effects can be computed by assuming that the motion is largely parallel to the camera image plane.

5.4 Capabilities

The previous section presented a model to predict the performance of the MoViT system. This section illustrates how such a model can be used to predict performance. Three use cases are covered. 1) performance in a specific workspace, 2) comparing marker sizes, 3) performance as worn on a limb.

All use cases in this section are based on the camera calibration parameters of a Logitech C920 webcam at a resolution of 1920×1080. The camera parameters were obtained using the camera calibration utility in OpenCV and are quoted in Table 4.1. Note that other resolutions and even other cameras may be used, so long as the camera calibration parameters are computed and provided to the model.

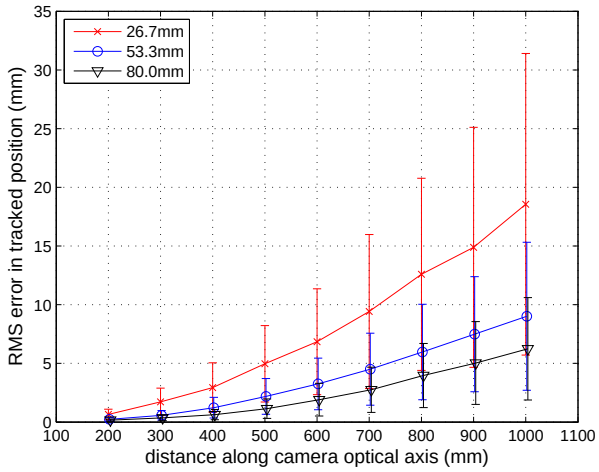


Figure 5.5: Using the model to compare different marker sizes. Tracking accuracy was predicted using markers sized 26.7×26.7, 53.3×53.3, and 80×80 mm. Each point represents mean (n=1000). Error bars indicate one standard deviation.

5.4.1 Tracking Accuracy in a Specific Workspace

This is a use case where position tracking accuracy is to be predicted for a specific workspace, given a marker size and a camera. In this experiment, a marker of size $x = y = s = 26.7$ mm ($q = 3.33$ mm) was placed at different locations along a horizontal plane at the camera level, at 100 mm intervals. The plane of the marker was held parallel to the camera image plane. Fig. 5.4 illustrates the RMS position tracking error at each location.

5.4.2 Comparing Markers of Different Sizes

This is a scenario where performance is to be compared for different marker sizes at different distances. The results are illustrated in Fig. 5.5, which shows the RMS position tracking error. In this experiment, markers sizes ($x = y = s$) used were 26.7 mm, 53.3 mm, and 80 mm ($q = 3.33$ mm, 6.67 mm, and 10.0 mm).

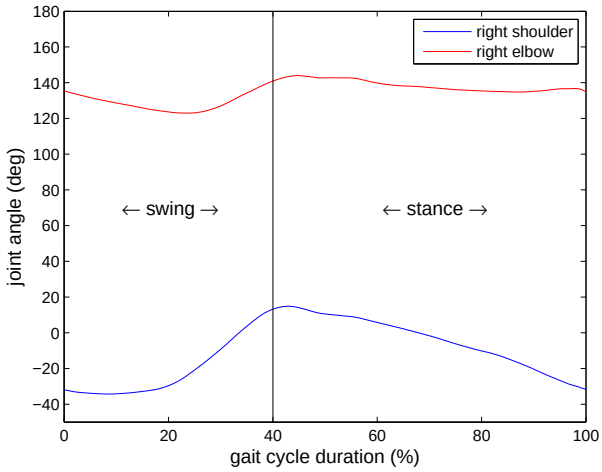


Figure 5.6: Joint angle data used from Righetti et al. [72] to simulate a marker placed on the wrist of a crawling infant.

Markers were placed at different locations along the optical axis (same as the Z axis in Fig. 5.4) from 200 mm to 1000 mm at 100 mm intervals. This scenario is identical to the experiment illustrated in Fig. 4.5a, with zero planar angle.

5.4.3 Marker Performance for a Specified Limb Motion

This is potentially the most helpful use case in the context of motion capture. This is used to predict the performance for a marker worn on a moving limb. A moving arm was simulated using the 2-link kinematic arm model. Joint angle data for the right arm of a crawling infant from Righetti et al. [72] were used (age group: 9-11 months). The joint angle data are illustrated in Fig. 5.6. Since these data were from an infant aged 9-11 months, size data for the 9-11 month age group were used from Table 5.2. Note that only the two flexion/extension angles (θ_1 and θ_4) are available, so arm motion is in a single plane. The marker size used was $x = y = s = 26.7$ mm ($q = 3.33$ mm).

The arm was rotated about the Y-axis so that the plane of motion of the

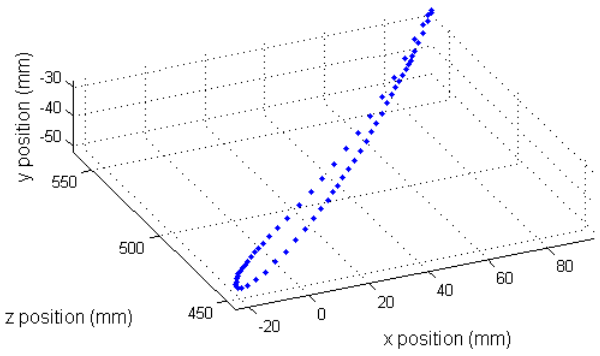


Figure 5.7: Simulated trajectory of the center of a marker placed on the wrist of a crawling infant. Camera at (0,0,0) mm.

arm was at 45 degrees relative to the camera image plane. Fig. 5.7 illustrates the motion of the center of the marker relative to the camera. To estimate instantaneous joint speeds, timing data were used from Righetti et al. [72]. The tracking accuracy of the marker position through the crawling motion is illustrated in Fig. 5.8.

Another aspect of the model is predicting the limit of the camera shutter exposure time. Fig. 5.10 shows the limiting shutter exposure time through the duration of the crawling motion. This is based on the instantaneous marker speed (see Fig. 5.9).

Note that the angles and positions at the start and end of the crawl cycle are approximately the same (Fig. 5.6, Fig. 5.7). But the speed at the start and end of the crawl cycle are not as close to each other (Fig. 5.9). One reason for this is that although the measured crawling motion was periodic in terms of angle and position, it was not as consistently periodic in time. The initial limb speed at the start of one cycle was typically different from the initial limb speed

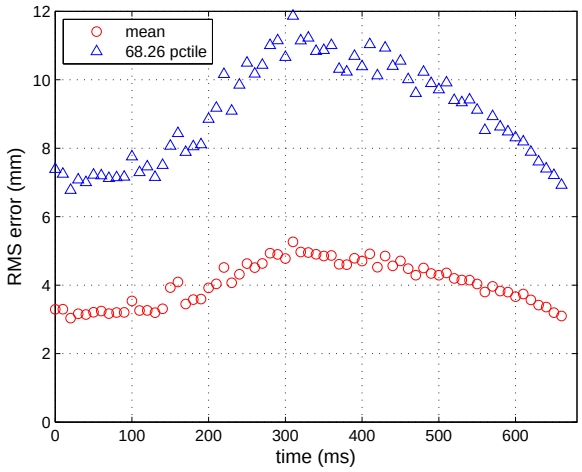


Figure 5.8: Using the MoViT model to predict tracking accuracy of a marker placed on a crawling infant's wrist. Points indicate mean error (n=1000). Error bars indicate 68.26 percentile. Based on sizing data from Table 5.2 and crawling data of the right arm of an infant from Righetti et al. [72]. Simulated at 100 Hz.

at the start of the very next cycle. This is illustrated for typical trajectories in the source for these data, i.e. Righetti et al. [72].

5.5 Comparison with Experimental Results

In the previous section, a model to predict the performance of the MoViT system along with some use cases was presented. In this section, a comparison of the model with experimental results is presented. The model predicts tracked position error as well as maximum shutter exposure time. Tracked position error is compared in this section.

5.5.1 Accuracy for Different Marker Sizes

The use case of comparing markers of 3 different sizes (Section 5.4.2) was examined. Results from the simulated setup of Section 5.4.2 were compared with the

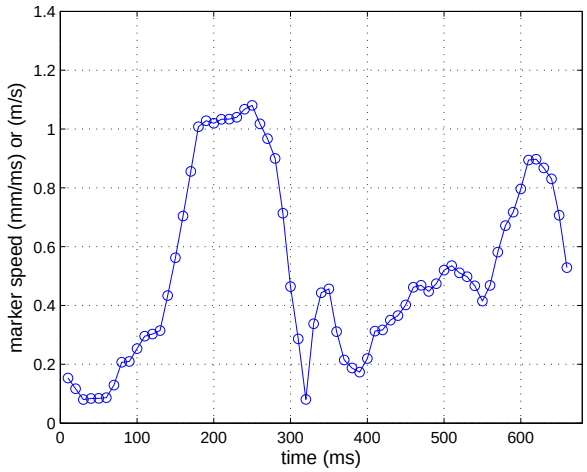


Figure 5.9: Using the MoViT model to estimate the speed of a marker placed on a crawling infant's wrist. Start and end speeds are not the same because infant speed typically varies from one cycle to the next.

results from the physical setup in Section 4.2.3. This was the setup illustrated in Fig. 4.5a where 3 different marker sizes were placed at different distances along the camera optical axis. In both cases, the same camera, i.e., the Logitech C920 (1920×1080 pixels) was used. The marker distance in the simulated setup was from 100 mm to 2000 mm. The marker distance in the physical setup was from 200 mm to 1000 mm. Only the common data points, i.e., from 200 mm through 1000 mm were used. Similarly, the simulated setup used only a planar angle of zero as compared to the planar angle values of 0, 30, and 60 degrees in the physical setup. Only the data points with planar angle 0 degree were thus used for comparison. The predicted RMS error (model) and the actual RMS error (physical setup) in tracked position error (%) were compared.

Fig. 5.11 shows a scatter plot of experimentally determined error versus model predicted error, for all the different marker sizes and poses involved. The correlation between the two can be observed from a linear fit to the data. From the linear fit, the relation between the actual and predicted error can be

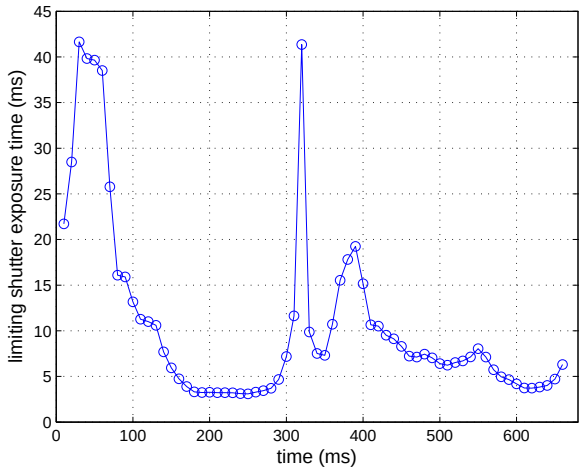


Figure 5.10: Using the MoViT model to estimate the maximum shutter exposure time required to capture a marker placed on a crawling infant's wrist at each instant.

estimated by:

$$error_{actual} = 1.50(error_{predicted}) + 0.0143 \tag{5.9}$$

The coefficient of determination (R^2) of the linear fit is 0.644. According to the linear fit, the RMS error that was measured experimentally is 1.5 times the error predicted by the model. The model predicted error is the same order of magnitude as the experimentally measured error. This means that a 3 % worst-case error would be predicted by the model as being 2 %. In other words, an error 15 mm would be predicted by the model as being 10 mm. For very small errors, the model seems to predict higher than experimental errors (pessimistic). For larger errors (greater than about 0.75 %) the model seems to predict lower than experimental errors (optimistic).

In general, the model predicted error is 33 % less than the actual error, i.e., the model underestimates the error. Therefore, to make the model more

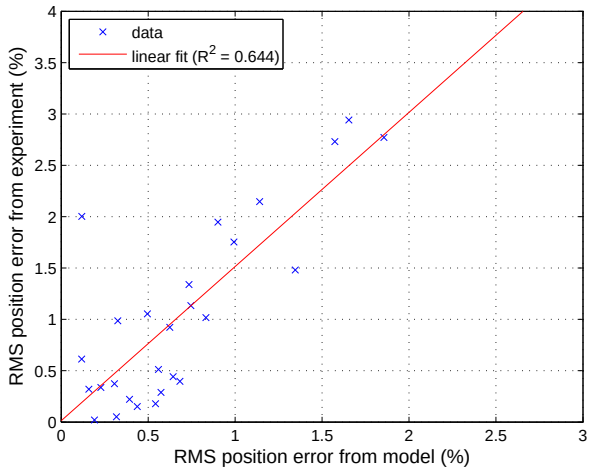


Figure 5.11: Comparison of mean error predicted by the model with the mean error measured from experiments.

realistic, either some elements of the model need to be improved, or additional elements need to be incorporated. One element that can be added is the behavior of the edge detection algorithms under different lighting and color balance conditions. These algorithms detect the edges and corners of a marker in an image. A bi-modal correction may also be used to improve the model.

5.6 Conclusion

This chapter presents a performance model for the MoViT system which uses square shaped AR markers for tracking. The model can be used for a marker worn on a limb, or for an explicitly defined set of marker poses. Marker size, camera parameters, and limb size can be defined as desired. Three use cases are highlighted to illustrate the capabilities of the model. Finally, some of the model error predictions are compared to error obtained from a physical test setup. The comparison of the two indicates that the error predicted by the model is the same order of magnitude as the error measured from physical tests.

While this is promising, there is room for improvement to make the model more realistic. Improvements can be made by refining existing elements of the model, or by including other elements not currently considered, or by combining all the unkown errors into a bi-modal or tri-modal correction.

Chapter 6

Conclusions and Future Work

6.1 Summary

The work in this dissertation is a subset of a larger effort to develop a new method of intervention for crawling-age infants with motor disorders such as Cerebral Palsy (CP). The fundamental idea of the intervention is that the infant is placed in the prone position on a robot (called SIPPC-3). The SIPPC-3 robot tracks the movements made by the infant. When crawling-like movements are detected, the robot physically moves the infant's body in that direction. This intervention can help with cognitive, as well as muscle development.

In this context, the latest generation of prototypes for the above intervention uses MEMS-based IMUs for capturing crawling motions. This is a common approach in motion capture. However, a limitation of this system is that MEMS IMUs use magnetometers. Since they are sensitive enough to measure the earth's relatively weak magnetic field, they are also sensitive to distortions in this field in the presence of large ferrous metals, or wires carrying mains electricity. This electromagnetic interference introduces significant errors in the

estimated body pose. This is compounded by the fact that sources of such interference are often hidden from view, and it is challenging to predict the onset of such interference in realtime.

In this context, there has been a need to develop a system that is not subject to the limitations of electromagnetic interference. This context drove the two research questions for this dissertation: 1) What are the requirements for capturing crawling motions of infants? 2) Can[14] a mocap system be developed that meets these requirements and is not subject to electromagnetic interference (EMI)?

To that end, the requirements for capturing infant crawling motion were compiled (ages 4-11 months). There is a severe lack of quantitative data on infant crawling motions. Only one source of literature with joint angle trajectories of 9-11 month old infants was located. Two other sources of data used to compile crawling motions were past experiments with the SIPPC-3 robot, and feedback from an expert on crawling development. Much of the data were available in the form of joint angle trajectories. To convert these into 3D Cartesian coordinate trajectories, forward kinematics using an infant kinematic skeleton were performed. The kinematic skeleton was based on infant sizing information. Much of the sizing information was not available directly and had to be derived. The end result of this effort was a data set of wrist and ankle motions was compiled for infants learning how to crawl. The smallest crawling motion that infants make is when they move a hand right or left along the floor to turn the body sideways. Considering the wrist, the start and end points of this motion are 74.6 mm apart. Motion speed is at least 42.7 mm/s. At 304 mm/s (95th percentile), infants move their wrists faster than their ankles. That is to

[14]This is a question of what is possible.

say, when considering 95th percentile speeds, the fastest motion that infants make is with the wrist, moving at 304 mm/s.

Once the requirements were compiled, the MoViT system featuring a vision-based marker tracking was developed. Initially, the approach was based on tracking colored markers. But after experiencing limitations of this approach, a planar pattern marker approach was developed. While typical marker-based approaches require multiple cameras, a planar pattern marker can be tracked in 3D by using a single camera.

The design of the MoViT system features planar marker bracelets worn on the wrists and ankles, and at least one reference marker worn at the back (Fig. 3.10). Tests were performed to evaluate the system. These included tracking a single marker at different positions and orientations and tracking one marker with reference to a reference marker. From the test with single markers, the worst-case error within 500 mm of the marker was measured to be less than 4 % (15 mm). For the test with the reference marker, the error turned out to be approximately 5 % (19.9 mm).

A model was developed to predict the performance of the MoViT system. This makes it easier to quickly predict performance by changing design parameters, such as camera resolution, marker size, distance, angle, etc. The most useful feature in the model is the ability to define a customizable infant limb model wearing a marker. By using joint angle trajectories, the model can be used to predict performance throughout the trajectory. Comparison of some of the predictions of the model with the physical setup indicated that the actual error was the same order of magnitude as the model predicted error. As a first step in refinement, the model can be improved by applying correction over the range of data collected.

6.2 Addressing the Research Questions

The first research question was: What are the requirements for capturing crawling motions of infants? The answer to that is the smallest significant motion that infants make as they learn how to crawl spans 74.6 mm. The tracking accuracy required is at least 37.3 mm. Motion should be captured at speeds of 42.7 mm/s to 304 mm/s.

The second research question was: To what extent does a mocap system, not subject to electromagnetic interference (EMI), meet the above requirements? The answer to that is as follows: with a worst-case error of about 19.9 mm, crawling motions of 74.6 mm or greater can potentially be detected. However, more evaluation needs to be done to determine whether this is practical for existing gesture recognition and filtering methods.

6.3 Future Work

Some more aspects of the MoViT system remain to be explored. Tracking error tests based on the camera field of view have been done in the wider horizontal field of view. For completeness, the narrower vertical field of view can also the tested. For this research, the emphasis has been on the 3D position tracking accuracy for the marker. More work needs to be done in testing for and modeling the 3D orientation tracking accuracy of the marker. Although the reference marker test has implicitly covered it, the combined error based on a reference marker as well as the target marker has not been thoroughly investigated. Perhaps this should be incorporated into the model. Finally, several tests with real infants would also help in evaluating the system. This would be most helpful

in determining the ideal number of cameras required to capture infant motion. This is probably best approached through empirical testing.

This work provides a limited insight into blurring and how it affects tracking. More work needs to be done to understand and model the effects of blurring. This includes two types of blurring. The first is motion blur. The value of the γ factor defined previously, which relates motion blur and the "speed limit" of a marker, must be determined. γ likely depends on a number of variables, e.g., white balance, type of edge detection algorithm used, and CCD properties. The effects of these variables must be determined and added to the model. The other type of blurring is one that takes place at very short shutter exposure times. At short exposure times, fewer photons strike the camera CCD. This reduces the signal-to-noise ratio, introduces a different type of blurring, and likely introduces position tracking errors. This type of blurring must also be investigated and modeled.

The system also needs to be packaged for field use, i.e., implemented in a portable form factor. Currently it is implemented on a laptop with an external webcam. This should be ported over to a platform like a Raspberry Pi 3 or a cell phone application. The MoViT system should have a reliable way of determining when a marker is too fast (blurred) to be detected. Machine vision algorithms are available to continue tracking objects that have suddenly become blurred (e.g., Bonnet et al. [15]). Perhaps that is an approach that can be implemented here. Finally, the mechanical design for the marker bracelet needs to be determined. It should be low cost, the material should be comfortable to wear yet rigid, it should be lightweight, and should be easy to put on.

Packaging for field use also involves deploying multiple cameras to ensure that all the MoViT markers on an infant's body are tracked. Most likely, at

least one camera will be required for target markers on each limb. Initially, two cameras could be attached to each leg of the SIPPC-3 robot frame. This number can be increased or decreased after testing. Since each camera will likely capture only a portion of the body, a number of reference markers on the body and the SIPPC-3 robot frame would be required for coordinate transforms to a common reference frame. One issue with this is that position and orientation errors from the various reference markers will begin accumulating. These errors could be reduced by simultaneously using multiple reference markers for each camera. Another issue is that if a marker, e.g., on the wrist, is visible to multiple cameras, then sensor fusion must be performed to combine the position tracking data from multiple cameras. While sensor fusion is not a problem in itself, each camera has the potential to increase or decrease the position tracking accuracy. Accuracy can increase because the same marker is visible from multiple cameras. But the opposite can also happen. Accuracy can decrease if some of the cameras have large tracking errors, which would propagate to the results of the sensor fusion process. This can be mitigated by developing metrics to continuously monitor the reliability of the tracked position data for every visible marker for each camera. This way, less weight will be given to cameras with less reliable readings at any given instant in the fusion process.

The MoViT system has one major limitation: occlusion. This is when a marker goes out of view and cannot be tracked by the system. If this happens for short periods of time, it can be addressed by fusing the camera data with IMU data. This is a well-established research area. Fusion of the two can be done in several ways, e.g., a particle filter-based approach (Tao et al. [86]), a complementary Kalman filter-based approach (Roetenberg et al. [74]) and a multi-rate extended Kalman filter approach (Hol et al. [36]).

While MEMS IMUs traditionally use magnetometers in their sensor fusion algorithms (e.g., Bosch Sensortec BNO055 [76]), some of them use only gyroscopes and accelerometers to estimate 3D orientation relative to an initial orientation. Examples of such sensors are the InvenSense MPU-6050 and MPU-9250 ([37], [38]). These should be tested and if the drift is acceptable, then they could be used for sensor fusion.

A slightly less noticeable issue with the MoViT system is that sometimes, a marker may not be detected due to extreme blurring effects, or large changes in automatic white balancing. This results in dropped frames, which drops the overall sampling rate of the visual system. Again, this issue can be resolved by fusing with data from an IMU, which can provide tracking information for the missing frames. Bright lighting can also be used.

The MoViT system does not have a very high sampling rate. For the webcam in Fig. 4.4 the sampling rate was about 15 Hz. For comparison, Southerland's IMU-based system [80] captures kinematics pose at 50 Hz. Although this reduced frame rate is partially due to processing limitations and dropped frames, a large part of it is due to camera shutter exposure time limitations. Tests in indoor lighting have shown automatically adjusting shutter exposure times to be several tens of milliseconds. There are several ways to increase the frame rate. One solution is to use a camera where the shutter exposure time can be manually controlled. This will increase the frame rate. The MoViT system is already in the process of being switched over over to the OV5647 camera for a Raspberry Pi 3. This camera allows manual control of the shutter exposure time. For cameras with automatic shutter exposure time control, more light can be used, which will force reduced camera shutter exposure times. Another solution is to use cameras with more sensitive CCD arrays (image sensors). Fi-

nally, as mentioned above, using sensor fusion with an IMU-based system, e.g., Southerland's suit system, can help increase the overall frame rate.

Bibliography

[1] 2016 Rice Baby Race. https://www.youtube.com/watch?v=dP1Dr7BCG2Q. Accessed February 2017.

[2] Advanced Realtime Tracking ARTTRACK5. https://ar-tracking.com/products/tracking-systems/arttrack5/. Accessed January 2018.

[3] Advanced Realtime Tracking ARTTRACK5/C. https://ar-tracking.com/products/tracking-systems/arttrack5c/. Accessed January 2018.

[4] Advanced Realtime Tracking SMARTTRACK. https://ar-tracking.com/products/tracking-systems/smarttrack/. Accessed May 2017.

[5] Advanced Realtime Tracking TRACKPACK/E. https://ar-tracking.com/products/tracking-systems/trackpacke/. Accessed January 2018.

[6] Cerebral Palsy Foundation Statistics. http://yourcpf.org/statistics. Accessed March 2017.

[7] The WHO Child Growth Standards. http://www.who.int/childgrowth/standards/en/. Accessed September 2017.

[8] Vicon. http://www.vicon.com/. Accessed March 2017.

[9] Vincent Agnus, Stéphane Nicolau, and Luc Soler. Illumination Independent and Accurate Marker Tracking Using Cross-Ratio Invariance. *IEEE Computer Graphics and Applications*, 35(5):22–33, 2015.

[10] Homayoun Bagherinia and Robert Manduchi. Robust real-time detection of multi-color markers on a cell phone. *Journal of Real-Time Image Processing*, 8(207):207–223, 2013.

[11] Bennett I. Bertenthal, Joseph J. Campos, and Rosanne Kermoian. An Epigenetic Perspective on the Development of Self-Produced Locomotion and Its Consequences. *Current Directions in Psychological Science*, 3(5):140–145, 1994.

[12] BTS Bioengineering. SMART-DX. http://www.arrayamed.com/fullaccess/product62file1.pdf. Accessed May 2017.

[13] 3D Bloom. Motion Capture Evaluative Essay. http://3dbloom.blogspot.com/2013/10/my-motion-capture-system-essay-brief.html. Accessed May 2017.

[14] Berta Bobath. The Very Early Treatment of Cerebral Palsy. *Developmental Medicine & Child Neurology*, 9(4):373–390, 1967.

[15] Vincent Bonnet, Nahema Sylla, Andrea Cherubini, Alejandro Gonzáles, Chirstine Azevedo Coste, Philippe Fraisse, and Gentiane Venture. Toward an Affordable and User-Friendly Visual Motion Capture System. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014.

[16] Andreas Breitenmoser, Laurent Kneip, and Roland Siegwart. A Monocular Vision-based System for 6D Relative Robot Localization. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 79–85, San Francisco, CA, USA, 2011.

[17] Suzann K. Campbell, Thubi H.A. Kolobe, Elizabeth T. Osten, Maureen Lenke, and Gay L. Girolami. Construct Validity of the Test of Infant Motor Performance. *Physical Therapy*, 75(7):585–596, 1995.

[18] Suzann K. Campbell, Thubi H.A. Kolobe, Benjamin D. Wright, and John M. Linacre. Validity of the Test of Infant Motor Performance for prediction of 6-, 9-and 12-month scores on the Alberta Infant Motor Scale. *Developmental Medicine & Child Neurology*, 44(04):263–272, 2002.

[19] Xi Chen, Sherry Liang, Stephen Dolph, Christina B. Ragonesi, James C. Galloway, and Sunil K. Agrawal. Design of a Novel Mobility Interface for Infants on a Mobile Robot by Kicking. *ASME Journal of Medical Devices*, 4(3):031006–1–031006–5, 2010.

[20] CMUcam. CMUcam5 Pixy. http://cmucam.org/projects/cmucam5. Accessed February 2018.

[21] Codamotion. Motion Capture and Movement Analysis Systems. http://www.codamotion.com/uploads/files/Codamotion%20RTE%20v2.pdf. Accessed October 2014.

[22] Benoît Delachaux, Julien Rebetez, Andres Perez-Uribe, and Héctor Fabio Satizábal Mejia. Indoor Activity Recognition by Combining One-vs.-All Neural Network Classifiers Exploiting Wearable and Depth Sensors.

In Ignacio Rojas, Gonzalo Joya, and Joan Cabestany, editors, *Advances in Computational Intelligence*, pages 216 – 223, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[23] Gaetano C. La Delfa, Salvatore Monteleone, Vincenzo Catania, Juan F. De Paz, and Javier Bajo. Performance Analysis of Visual Markers for Indoor Navigation Systems. *Frontiers of Information Technology & Electronic Engineering*, 17(8):730–740, 2016.

[24] Vision Doctor. Camera calculator - Calculating the shutter exposure time. `http://www.vision-doctor.com/en/camera-calculations/calculation-exposure-time.html`. Accessed January 2018.

[25] Diego Brito dos Santos Cesar, Christopher Gaudig, Martin Fritsche, Marco A. dos Reis, and Frank Kirchner. An Evaluation of Artificial Fiducial Markers in Underwater Environments. In *OCEANS'15 MTS/IEEE Genova*, 2015. Online.

[26] Linda Fetters, Inbal Sapir, Yu-Ping Chen, Masayoshi Kubo, and Ed Tronick. Spontaneous Kicking in Full-Term and Preterm Infants With and Without White Matter Disorder. *Developmental Psychobiology*, 52(6):524–536, 2010.

[27] Mark Fiala. ARTag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 590–596, 2005.

[28] Mark Fiala. Designing Highly Reliable Fiducial Markers. *IEEE Transactions on Pattern analysis and machine intelligence*, 32(7):1317–1324, 2010.

[29] Centers for Disease Control and Prevention. Data & Statistics for Cerebral Palsy. `https://www.cdc.gov/ncbddd/cp/data.html`. Accessed February 2018.

[30] Centers for Disease Control and Prevention. National Center for Health Statistics Clinical Growth Charts. `https://www.cdc.gov/growthcharts/clinical_charts.htm`. Accessed September 2017.

[31] Centers for Disease Control and Prevention. Screening and Diagnosis of Cerebral Palsy. `http://www.cdc.gov/ncbddd/cp/diagnosis.html`. Accessed March 2015.

[32] Centers for Disease Control and Prevention. Anthropometric Reference Data for Children and Adults: United States, 2011-2014. `https://www.cdc.gov/nchs/data/series/sr_03/sr03_039.pdf`, August 2016. Accessed September 2017.

[33] Robert L. Freedland and Bennett I. Bertenthal. Developmental Changes in Interlimb Coordination: Transition to Hands-and-Knees Crawling. *Psychological Science*, 5(1):26–32, 1994.

[34] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. Automatic Generation and Detection of Highly Reliable Fiducial Markers Under Occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014.

[35] Mustafa A. Ghazi, Michael D. Nash, Andrew H. Fagg, Lei Ding, Thubi H. A. Kolobe, and David P. Miller. *Novel Assistive Device for Teaching Crawling Skills to Infants*, pages 593–605. Springer International Publishing, 2016.

[36] Jeroen D. Hol, Thomas B. Schön, Henk Luinge, Per J. Slycke, and Fredrik Gustafsson. Robust Real-Time Tracking by Fusing Measurements from Inertial and Vision Sensors. *Journal of Real-Time Image Processing*, 2(2-3):149–160, November 2007.

[37] InvenSense. MPU-6050 Six-Axis (Gyro + Accelerometer) MEMS MotionTracking™Devices. `https://www.invensense.com/products/motion-tracking/6-axis/mpu-6050/`. Accessed January 2018.

[38] InvenSense. MPU-9250 Nine-Axis (Gyro + Accelerometer + Compass) MEMS MotionTracking™Device. `https://www.invensense.com/products/motion-tracking/9-axis/mpu-9250/`. Accessed January 2018.

[39] Suh-Fang Jeng, Chen Li-Chiou, and Kuo-Inn Tsou Yao. Kinematic Analysis of Kicking Movements in Preterm Infants With Very Low Birth Weight and Full-Term Infants. *Physical Therapy*, 82(2):148–159, 2002.

[40] Leonard W. Wilson Jr. Correction of Kinematic Data from a Self-Initiated Prone Position Crawler Trainer for Infants with Cerebral Palsy. Master's thesis, School of Biomedical Engineering, University of Oklahoma, 2017.

[41] Hirokazu Kato and Mark Billinghurst. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, CA, USA, 1999.

[42] Thubi H. A. Kolobe. Personal interview, 2017. College of Allied Health, University of Oklahoma Health Sciences Center, Oklahoma City.

[43] Kenneth Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

[44] Alberto López-Cerón and José Mara Cañas. Accuracy Analysis of Marker-Based 3D Visual Localization. XXXVII Jornadas de Automática Workshop, 2016.

[45] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, and C. Disselhorst-Klug. Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy. *Human Movement Science*, 25(2):125–144, 2006.

[46] Jonathan Meyer. A Low Cost, Vision Based Micro Helicopter System for Education and Control Experiments. Master's thesis, School of Aerospace and Mechanical Engineering, University of Oklahoma, 2014.

[47] David P. Miller, Andrew H. Fagg, Lei Ding, Thubi H.A. Kolobe, and Mustafa A. Ghazi. Robotic Crawling Assistance for Infants with Cerebral Palsy. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 36–38, 2015.

[48] David P. Miller, Anne Wright, Randy Sargent, Rob Cohen, and Teresa Hunt. Attitude and Position Control Using Real-Time Color Tracking. In *Proceedings of the AAAI-97/IAAI-97 Conference*, pages 1026–1031, Providence, RI, USA, July 1997.

[49] Organic Motion. OpenStage-Markerless Mocap for Education. `http://www.organicmotion.com/mocap-for-education/`.

[50] Myagmarbayar Nergui, Yuki Yoshida, Nevrez Imamoglu, Jose Gonzalez, Masashi Sekine, and Wenwei Yu. Human Motion Tracking and REcognition using HMM by a Mobile Robot. *International Journal of Intelligent Unmanned Systems*, 1(1):76–92, 2013.

[51] Mikkel D. Olsen, Anna Herskind, Jens B. Nielsen, and Rasmus R. Paulsen. Body Part Tracking of Infants. In *2014 22nd International Conference on Pattern Recognition*, pages 2167–2172, 2014.

[52] Mikkel D. Olsen, Anna Herskind, Jens B. Nielsen, and Rasmus R. Paulsen. Model-based Motion Tracking of Infants. In *Computer Vision-ECCV 2014 Workshops*, pages 673–685. Springer, 2014.

[53] Edwin Olson. AprilTag: A Robust and Flexible Visual Fiducial System. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407. IEEE, May 2011.

[54] OpenCV. About. `http://opencv.org/about.html`. Accessed July 2017.

[55] OpenCV. ArUco marker detection (aruco module). http://docs.opencv.org/3.1.0/d9/d6d/tutorial_table_of_content_aruco.html. Accessed March 2017.

[56] OpenCV. Camera Calibration and 3D Reconstruction. http://docs.opencv.org/3.1.0/d9/d0c/group__calib3d.html. Accessed July 2017.

[57] Optitrack. Flex 13. http://www.optitrack.com/products/flex-13/. Accessed May 2017.

[58] Optitrack. Flex 3. http://www.optitrack.com/products/flex-3/. Accessed May 2017.

[59] Optitrack. Prime 13. http://www.optitrack.com/products/prime-13/. Accessed May 2017.

[60] Optitrack. Prime 13W. http://www.optitrack.com/products/prime-13w/. Accessed May 2017.

[61] Optitrack. Prime 17W. http://www.optitrack.com/products/prime-17w/. Accessed May 2017.

[62] Optitrack. Prime 41. https://www.optitrack.com/products/prime-41/. Accessed May 2017.

[63] Optitrack. Slim 13E. http://www.optitrack.com/products/slim-13e/. Accessed May 2017.

[64] Optitrack. Slim 3U. https://www.optitrack.com/products/slim-3u/. Accessed May 2017.

[65] Optitrack. V120 Duo. http://www.optitrack.com/products/v120-duo/. Accessed May 2017.

[66] Optitrack. V120 Trio. http://www.optitrack.com/products/v120-trio/. Accessed May 2017.

[67] Orbbec. Orbbec Astra, Astra S, & Astra Pro. https://orbbec3d.com/product-astra/. Accessed February 2016.

[68] Phasespace. Impulse X2 Motion Capture System. http://phasespace.com/impulse-motion-capture.html. Accessed May 2017.

[69] Peter E. Pidcoe and Hlapang A. Kolobe. Self Initiated Prone Progressive Crawler, 2015.

[70] Qualisys. Miqus. http://www.qualisys.com/cameras/miqus/. Accessed May 2017.

[71] Qualisys. Oqus. http://www.qualisys.com/cameras/oqus/. Accessed May 2017.

[72] Ludovic Righetti, Anna Nylén, Kerstin Rosander, and Auke Jan Ijspeert. Kinematic and Gait Similarities between Crawling Human Infants and Other Quadruped Mammals. *Frontiers in Neurology*, 6(17), 2015.

[73] Nelci Adriana Cicuto Ferreira Rocha, Fernanda Pereira dos Santos Silva, and Eloisa Tudella. The Impact of Object Size and Rigidity on Infant Reaching. *Infant Behavior and Development*, 29(2):251–261, 2006.

[74] Daniel Roetenberg and Peter H. Veltink. Camera-Marker and Inertial Sensor Fusion for Improved Motion Tracking. In *Gait & Posture*, volume 22, pages 1–53, 2005.

[75] Randy Sargent, Bill Bailey, Carl Witty, and Anne Wright. The importance of fast vision in winning the first micro-robot world cup soccer tournament. *Robotics and Autonomous Systems*, 21(2):139–147, 1997.

[76] Bosch Sensortec. BNO055. https://www.bosch-sensortec.com/bst/products/all_products/bno055. Accessed January 2018.

[77] Monique Shotande. Personal interview, 2017. School of Computer Science, University of Oklahoma, Norman.

[78] Beth A. Smith, Ivan A. Trujillo-Priego, Christianne J. Lane, James M. Finley, and Fay B. Horak. Daily Quantity of Infant Leg Movement: Wearable Sensor Algorithm and Relationship to Walking Onset. *Sensors*, 15(8):19006–19020, 2015.

[79] Richard G. Snyder, Lawrence W. Schneider, Clyde L. Owings, Herbert M. Reynolds, D. Henry Golomb, and M. Anthony Schork. Anthropometry of Infants, Children, and Youths to Age 18 for Product Safety Design. Technical Report UM-HSRI-77-17, Highway Safety Research Institute, Ann Arbor, MI, USA, 1977.

[80] Joshua B. Southerland. Activity Recognition and Crawling Assistance Using Multiple Inexpensive Inertial Measurement Units. Master's thesis, School of Computer Science, University of Oklahoma, 2012.

[81] Innovision Systems. Max100 - Twelve (12) Camera, 100/200 Hz Basic System. http://www.innovision-systems.com/Downloads/SystemsDetails/Max100_12.pdf. Accessed May 2017.

[82] Innovision Systems. Max300 - Twelve (12) Camera, 320 Hz High Speed System. http://www.innovision-systems.com/Downloads/SystemsDetails/Max300_12.pdf. Accessed May 2017.

[83] Innovision Systems. MaxPRO 3D - Four (4) Camera, 160/500 Hz Basic System. http://www.innovision-systems.com/Downloads/SystemsDetails/MP_160_4.pdf. Accessed May 2017.

[84] Innovision Systems. MaxTRAQ 3D - Four (4) Camera, 160/500 Hz Basic System. http://www.innovision-systems.com/Downloads/SystemsDetails/M3_160_4.pdf. Accessed May 2017.

[85] Innovision Systems. MaxTRAQ 3D - Four (4) Camera, 60/120 Hz Basic System. www.innovision-systems.com/Downloads/SystemsDetails/M3_60_4.pdf. Accessed May 2017.

[86] Yaqin Tao and Huosheng Hu. A Novel Sensing and Data Fusion System for 3-D Arm Motion Tracking in Telerehabilitation. *IEEE Transactions on Instrumentation and Measurement*, 57(5):1029–1040, May 2008.

[87] Phoenix Technologies. Visualeyez III VZ10K/VZ10K5. http://ptiphoenix.com/products/trackers/VZ10K_VZ10K5. Accessed May 2017.

[88] Phoenix Technologies. VZ4000v. http://ptiphoenix.com/products/trackers/VZ4000v. Accessed May 2017.

[89] Phoenix Technologies. VZ4050. http://ptiphoenix.com/products/trackers/VZ4050. Accessed May 2017.

[90] Vicon. Vero. https://www.vicon.com/products/camera-systems/vero. Accessed May 2017.

[91] Vicon. Vicon Vantage. https://www.vicon.com/products/camera-systems/vantage. Accessed May 2017.

[92] Robert Wang, Sylvain Paris, and Jovan Popović. Practical Color-based Motion Capture. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 139–146, New York, NY, USA, 2011. ACM.

[93] Tingfan Wu, Juan Artigas, Whitney Mattson, Paul Ruvolo, Javier Movellan, and Daniel Messinger. Collecting A Developmental Dataset of Reaching Behaviors: First Steps. In *IROS2011 Workshop on Cognitive Neuroscience Robotics*, pages 47–52, 2011.

[94] Xcitex. ProCapture High-Speed Multi-Camera Motion Capture System. http://www.xcitex.com/procapture-motion-capture-systems.php. Accessed May 2017.

[95] Qi L. Xiong, Xiao Y. Wu, Nong Xiao, Si Y. Zeng, Xiao P. Wan, Xiao L. Zheng, and Wen S. Hou. Antagonist muscle co-activation of limbs in human infant crawling: A pilot study. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2115–2118, 2015.

[96] Xsens. Xsens MVN: The Animator's Tool. `http://www.xsens.com/wp-content/uploads/2013/11/mvn-leaflet.pdf`.

[97] Xiang Zhang, Stephan Fronz, and Nassir Navab. Visual Marker Detection and Decoding in AR Systems: A Comparative Study. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 97–106, 2002.

[98] Zhengyou Zhang. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia*, 19(2):4–10, 2012.

[99] Huiyu Zhou and Huosheng Hu. Human motion tracking for rehabilitation-A survey. *Biomedical Signal Processing and Control*, 3(1):1–18, 2008.

Appendix A

Theoretically Estimated Error for Marker Position Tracking

In this appendix, a theoretical estimate of the position tracking error in the experiment in Section 4.2.3 is presented. In this experiment, the MoViT marker position was tracked using OpenCV, while marker position, planar angle and size were varied. The experimental results were presented in Section 4.2.4. This appendix provides a theoretical estimate of the error.

The position error is derived in mm per unit pixel error. This is because pixel discretization in the camera CCD introduces an error of up to 0.5 pixels. This error propagates through the marker pose estimation process. Furthermore, edge detection algorithms used to detect marker patterns in OpenCV can also introduce pixel error.

A.1 Defining a Marker in 3D

The 3D coordinates of a square-shaped marker are defined in terms of its geometry, a rotation, and a translation. The geometry of the marker is defined with its center at the origin, with all 4 points on the x-y plane (see Fig. 5.2). Its position in 3D is defined by a rotation transform and a translation transform. The rotation transform is a 3 $\times$ 3 matrix. The translation transform is a 3 $\times$ 1 matrix. Thus, the 3D coordinates for the 4 corners of a marker of size $L \times L$ are defined as:

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} -L/2 & L/2 & L/2 & -L/2 \\ L/2 & L/2 & -L/2 & -L/2 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}.$$

The ordering of the points above is clockwise, starting from the top left corner of the marker. For brevity, consider the top right corner only:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} L/2 \\ L/2 \\ 0 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}.$$

Note that by definition, the geometry of any of the four corners of a marker will always be defined at $[\pm\frac{L}{2} \ \pm\frac{L}{2} \ 0]^T$. The third coordinate is always zero. Therefore, by definition, row 3 of the rotation matrix is redundant. This simplifies

the geometry model to:

$$\begin{bmatrix} x \\ \\ y \\ \\ z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ \\ r_{21} & r_{22} \\ \\ r_{31} & r_{32} \end{bmatrix} \begin{bmatrix} L/2 \\ \\ L/2 \end{bmatrix} + \begin{bmatrix} t_1 \\ \\ t_2 \\ \\ t_3 \end{bmatrix}.$$

This implies that to define the 3D points of a square marker with points that are coplanar, the rotation matrix transform requires only the first two columns of any 3×3 rotation matrix.

A.2 Camera Model

The camera model projects a real world point $W = [x\ y\ z]^T$ (in mm) to the camera image plane $C = [u\ v]^T$ (in pixels) (see Fig. 5.2). It is based on the pinhole camera model modified for lens distortion effects. Given:

$$x' = x/z, \quad y' = y/z, \tag{A.1}$$

the radial and tangential distortions can be approximated by:

$$x'' = x'(1 + k_1 r^2 + k_2 r^4) + 2p_1 x' y' + p_2(r^2 + 2x'^2), \tag{A.2}$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2y'^2) + 2p_2 x' y', \tag{A.3}$$

where $r^2 = x'^2 + y'^2$. k_1 and k_2 are radial distortion coefficients, and p_1 and

p_2 are tangential distortion coefficients. More details on lens distortion can be found in the OpenCV documentation for camera calibration [56]. Finally, the pixel coordinates $C = [u\ v]^T$ can be computed by:

$$u = f_x x'' + c_x, \tag{A.4}$$

$$v = f_y y'' + c_y. \tag{A.5}$$

A.3 Setting Up the Pose Estimation Problem

The above set of equations represent a perspective n-point problem (PNP), with 4 known points and 6 unknown variables. Of these, 3 unknowns define the position of the marker, i.e. $[t_1\ t_2\ t_3]^T$. The other 3 define the 3D rotation. Although the rotation transform above shows 6 unknowns, a 3×3 rotation matrix can be defined in terms of 3 rotation angles.

The error in each coordinate of the tracked position (i.e., t_1, t_2, t_3) is sensitive to the error in pixel coordinates (i.e., u and v). Thus, for each position coordinate, the error per unit pixel can be defined as $\frac{\partial t_1}{\partial u}$, $\frac{\partial t_2}{\partial u}$, $\frac{\partial t_3}{\partial u}$, $\frac{\partial t_1}{\partial v}$, $\frac{\partial t_2}{\partial v}$, and, $\frac{\partial t_3}{\partial v}$.

A.4 Simplifying Pose Estimation for a Specific Case

Consider the simplified case where a marker is placed at different positions in front of the camera, along the center-line, with a varying planar angle. This is

the experiment described in Section 4.2.3. In this case, only one rotation, i.e. about the y-axis, is present. The 3D coordinates of the top right corner are defined as:

$$\begin{bmatrix} x \\ \\ y \\ \\ z \end{bmatrix} = \begin{bmatrix} cos\phi & 0 \\ \\ 0 & 1 \\ \\ -sin\phi & 0 \end{bmatrix} \begin{bmatrix} L/2 \\ \\ L/2 \end{bmatrix} + \begin{bmatrix} t_1 \\ \\ t_2 \\ \\ t_3 \end{bmatrix}.$$

Alternatively,

$$x = \frac{L}{2} cos\phi + t_1, \tag{A.6}$$

$$y = \frac{L}{2} + t_2, \tag{A.7}$$

$$z = -\frac{L}{2} sin\phi + t_3. \tag{A.8}$$

Applying projection onto the camera image plane,

$$x' = x/z = \frac{2t_1 + Lcos\phi}{2t_3 - Lsin\phi}, \tag{A.9}$$

$$y' = y/z = \frac{L + 2t_2}{2t_3 - Lsin\phi}. \tag{A.10}$$

Assuming that the radial and tangential distortion effects are negligible, the pinhole camera model gives the pixel coordinates (u, v) on the camera image plane:

$$u = f_x x' + c_x = f_x \frac{2t_1 + Lcos\phi}{2t_3 - Lsin\phi} + c_x, \tag{A.11}$$

$$v = H - (f_y y' + c_y) = H - (f_y \frac{L + 2t_2}{2t_3 - Lsin\phi} + c_y). \tag{A.12}$$

This gives the following relationships for computing t_1, t_2, and t_3:

$$t_1 = \frac{c_x Lsin\phi - Lusin\phi - f_x Lcos\phi - t_3(-2u + 2c_x)}{2f_x}, \tag{A.13}$$

$$t_3 = \frac{c_x Lsin\phi - Lusin\phi - f_x Lcos\phi - 2f_x t_1}{-2u + 2c_x}, \tag{A.14}$$

$$t_2 = \frac{HLsin\phi - c_y Lsin\phi - Lvsin\phi + f_y L + t_3(2c_y - 2H + 2v)}{-2f_y}, \tag{A.15}$$

$$t_3 = \frac{HLsin\phi - c_y Lsin\phi - Lvsin\phi + f_y L + 2f_y t_2}{-(2c_y - 2H + 2v)}, \tag{A.16}$$

where f_x and f_y are the focal lengths, c_x and c_y are the principal points, and H is the vertical image resolution. H simply transforms the coordinates from the physical frame to the pixel frame where the physical frame has the $+Y$ axis pointing upwards. Note that for t_3, two different equations are available. These are two different ways of deriving t_3. Either one may be used for error estimation. Thus, the error terms in mm/pixel are given by the partial derivatives:

$$\frac{\partial t_1}{\partial u} = \frac{2t_3 - Lsin\phi}{2f_x}, \tag{A.17}$$

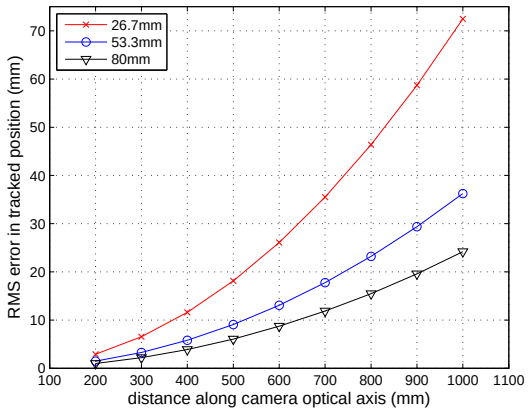
$$\frac{\partial t_2}{\partial v} = -\frac{2t_3 - Lsin\phi}{2f_y}, \tag{A.18}$$

$$\frac{\partial t_3}{\partial u} = \frac{-2(2f_x t_1 + f_x Lcos\phi - c_x Lsin\phi + Lusin\phi)}{(2c_x - 2u)^2} - \frac{Lsin\phi}{2c_x - 2u}. \tag{A.19}$$

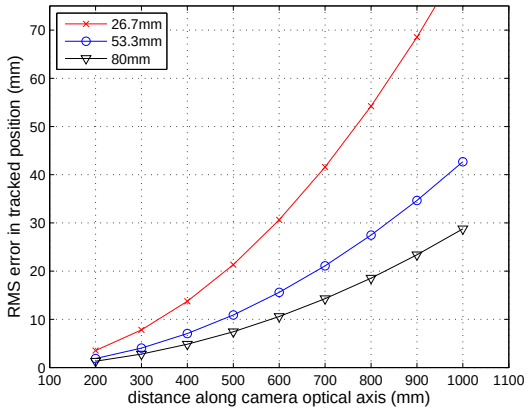
Physically, the above terms represent the position error in mm per unit pixel of error. That is to say, for the image of a point in 3D, if the point detection is off by one pixel, then the above terms indicate the resulting error in computing each of the 3D coordinates of that point (i.e., error in t_1, t_2, and t_3). The analogy for a marker with four corners is that the marker appears larger by one pixel on all sides.

A.5 Theoretically Estimated Error

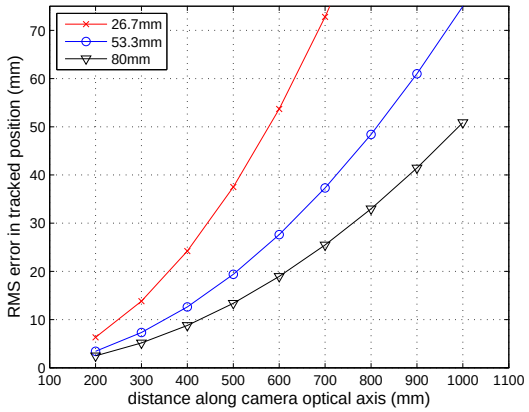
The RMS error in mm/pixel, for 3 different angles, is shown in Fig. A.1. Although RMS error includes all three errors $\frac{\partial t_1}{\partial u}$, $\frac{\partial t_2}{\partial v}$, and $\frac{\partial t_3}{\partial u}$, the dominant error by far is $\frac{\partial t_3}{\partial u}$. That is to say, mainly the Z coordinate (t_3) is affected by pixel errors. Fig. A.1 may be compared to Fig. 4.7. The error is larger in Fig. A.1, because Fig. A.1 represents position error based on an error of 1 pixel in the camera image. In contrast, Fig. 4.7 represents position error based on approximately 0.5 pixel error in the image.



(a) 0 degrees



(b) 30 degrees



(c) 60 degrees

Figure A.1: Theoretically estimated position tracking error. Absolute RMS position error at angles 0, 30, and 60 degrees. Error is based on (A.17), (A.18), and (A.19). Plots have been scaled to match Fig. 4.7 for comparison.

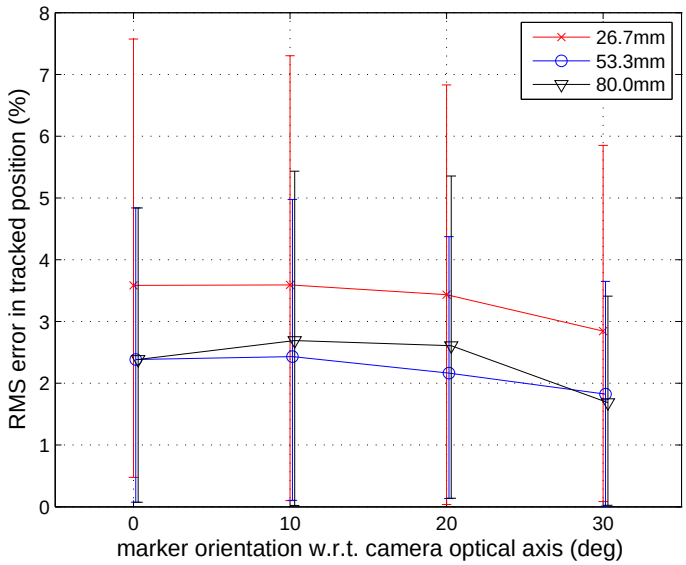
Appendix B

Supplementary Results: Varying Camera View Angle

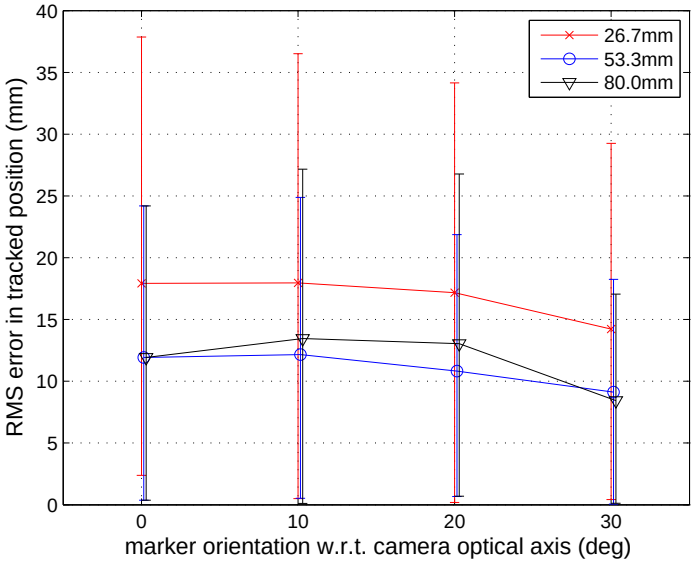
Supplementary results for the test in Section 4.2.5 are presented below. In this test, markers were rotated in the camera field of view, and the tracked position error was observed. For details of the setup, see Section 4.2.5.

The results below were obtained from testing with two more cameras. One of these cameras was a Logitech model C920, the same model as the one used in Section 4.2.5. The other camera was a Logitech model C615.

These cameras were tested to verify that the position tracking error is not a function of marker view angle in the camera field of view. When the results in Section 4.2.6 (Fig. 4.9) are viewed in isolation, it can be incorrectly concluded that the position tracking error increases with the marker view angle in the camera field of view. The collective results from all three cameras indicate that there is no correlation between tracking error and marker orientation with reference to camera optical axis. The results for the second Logitech C920 are presented in Fig. B.1, and for the Logitech C615 are presented in Fig. B.2.

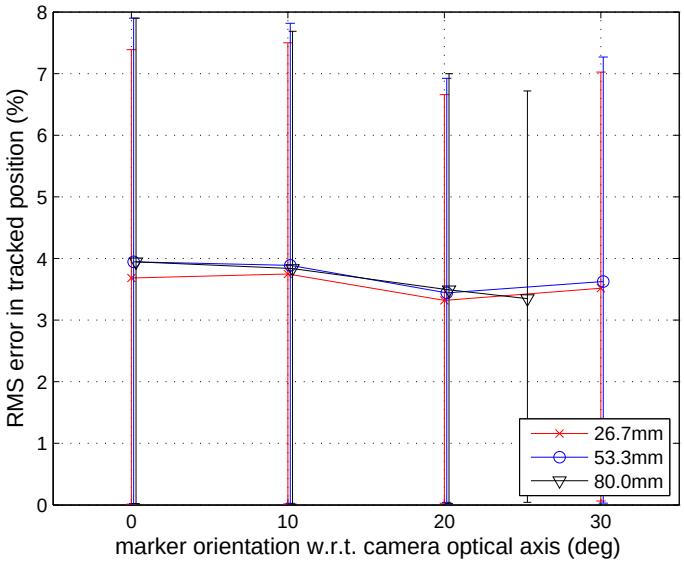


(a) percentage error

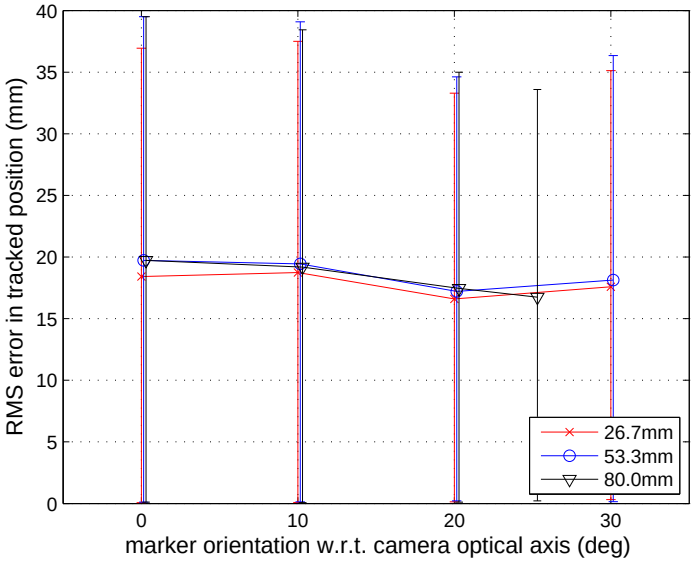


(b) absolute error

Figure B.1: Supplementary results for accuracy experiment 2 (camera view angle). A second camera, also a Logitech model C920, was used. Error variation as the marker is rotated away from the optical axis of the camera. Error bars indicate lower and upper quartiles. The radial distance to the marker is 500 mm. Sample size $n = 20$.



(a) percentage error



(b) absolute error

Figure B.2: Supplementary results for accuracy experiment 2 (camera view angle). A third camera, Logitech model C615, was used. Error variation as the marker is rotated away from the optical axis of the camera. Error bars indicate lower and upper quartiles. The radial distance to the marker is 500 mm. Sample size $n = 20$.

Appendix C

Supplementary Results: Orientation Tracking Error

This dissertation has been focused on position tracking error. Orientation tracking error can provide a useful insight when marker tracking involves multiple cameras and reference markers. To that end, this appendix provides supplementary results for the test in Section 4.2.3. While Section 4.2.4 presented position tracking error, this appendix provides orientation tracking error for the same test. An understanding of orientation tracking error is important for future work.

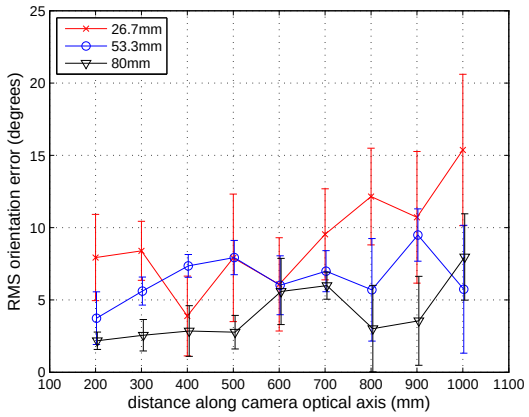
Most studies analyzing orientation error present plots for the three angles (roll, pitch, yaw) separately. Comparing such plots is not very intuitive. Therefore, for this application, a single orientation error metric is defined. It is analogous to RMS position error:

$$error_{orientation} = \sqrt{(error_{roll})^2 + (error_{pitch})^2 + (error_{yaw})^2}, \qquad \text{(C.1)}$$

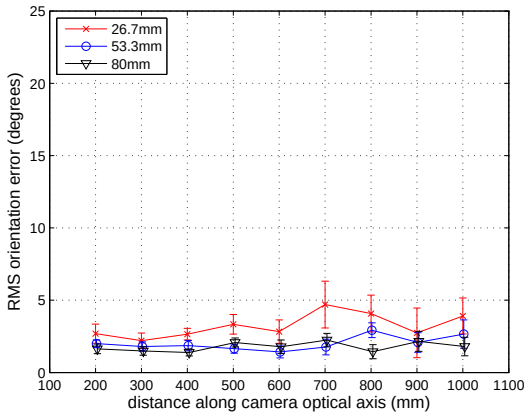
where $error_{roll}$ is the error in roll angle, $error_{pitch}$ is the error in pitch angle, and $error_{yaw}$ is the error in yaw angle. The results are shown in Fig. C.1 for 3 different marker sizes and 3 different planar angles.

The results show an interesting trend. The orientation error for the 80 mm marker is consistently small for all 3 planar angles, for distances up to 500 mm (relevant range for the SIPPC-3 robot). For marker sizes 26.7 mm and 53.3 mm, orientation estimates are worst at 0 degrees planar angle. They improve greatly for 30 degrees and 60 degrees, where the worst case orientation error is less than 4 degrees. This trend is the reverse of the results for position error (Section 4.2.4), where position error was best for 0 degrees planar angle. This trend in orientation error is expected due to the nature of perspective geometry. Perspective projection effects are more noticeable when there is at least some angle between the plane being projected (in this case the marker plane) and the camera image plane. When this angle is zero, perspective projection effects are not very clear. Hence, using perspective geometry principles to estimate orientation results in larger errors.

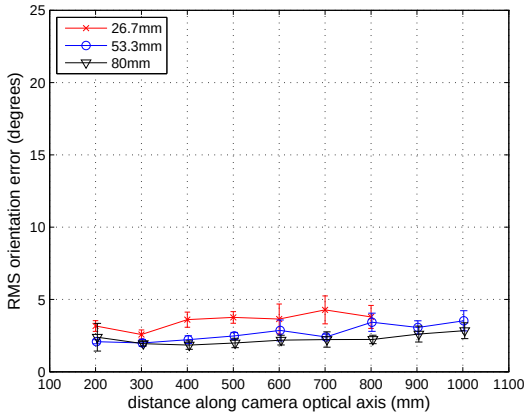
In terms of using reference markers for tracking, these results imply that if a reference marker is relatively small, then it must be oriented at an angle to the camera image plane so as to minimize the propagation of the orientation error. If the reference marker plane has to placed parallel to the camera image plane, then it should be large enough (larger than 53.3 mm) so that the orientation error remains small.



(a) 0 degrees



(b) 30 degrees



(c) 60 degrees

Figure C.1: Orientation tracking error. Absolute RMS angle error at planar angles 0, 30, and 60 degrees. Error bars indicate standard deviation. Marker sizes $x = y = 8q$ were 26.7 mm, 53.3 mm, and 80.0 mm. Sample size $n = 20$.